2021

An Assessment of the Impacts of Social Media Inputs and Court Case Information on Mitigating Insider Threats

Robert Jones
Nova Southeastern University, rj631@mynsu.nova.edu

An Assessment of the Impacts of Social Media Inputs and Court Case Information on Mitigating Insider Threats

by

Robert W. Jones

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Information Assurance

College of Computing and Engineering
Nova Southeastern University

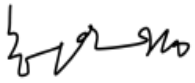
2021

Approval and Signature

We hereby certify that this dissertation, submitted by Robert Jones conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

1-22-2021

Ling Wang, Ph.D — **Date**
Chairperson of Dissertation Committee



1/22/2021

Inkyoung Hur, Ph.D. — **Date**
Dissertation Committee Member

1/22/2021

Junping Sun, Ph.D. — **Date**
Dissertation Committee Member

Approved:

1/22/2021

Meline Kevorkian, Ed.D. — **Date**
Dean, College of Computing and Engineering

College of Computing and Engineering
Nova Southeastern University

2021

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

An Assessment of the Impacts of Social Media Inputs and Court Case on Mitigating Insider Threats

by
Robert W. Jones
March 2021

The insider threat is a global problem that impacts organizations and produces a gamut of undesired outcomes. Businesses often experience lost revenue and stolen trade secrets, which can leave a tarnished reputation. Insider threats can also cause harm to individuals and national security.

Past efforts have not mitigated the problem in its entirety. Documented instances of insider threats are as recent as March 2020. Many researchers have focused on monitoring technologies and relying on human monitoring in a reactive posture. An ideal solution would scrutinize an individual's character and ascertain whether unique traits associated with actors of insider threats are apparent within the preemployment vetting process.

This study leveraged various input data streams and applied theory-driven behaviors that are associated with fraudulent activities. The research followed a Design Science Research (DSR) methodology to produce sentiment analysis of IT artifacts, and ranked individuals' level of trustworthiness, conducive within the hiring process.

Lab experiments were used to answer the research questions, provided valuable insight with fraudulent activities, and discovered commonalities with negative sentiments found in social media tweets. First, literature was defined and reviewed to address mitigation of insider threats in one form or another. Second, artifacts were from the sum of all data components; these artifacts proved to be informative during the construction within each lab experiment. Finally, the lab experiments provided helpful contributions to the study. For instance, across all lab experiments, common themes emerged from four negative sentiment scores. These scores were later illustrated under the S140-negScore, AFINN-negScore, SentiWordnet-negScore, and NRC-Hash-Sent-negScore. Behavioral theories did not always appear within each artifact; however, the routine activity theory was the most prevalent and was detailed in the lab experiments.

The research extends previous and relevant research, thus leveraging social inputs and fraudulent data extracted from the legal system as a foundation for a way forward. An insider threat can be mitigated through leveraging social media data.

Acknowledgments

First, I want to thank God for putting the dream in my heart and getting me through the process. I also have family and friends who offered encouragement along the way. I appreciate those who went before me, who had the willingness to answer questions in a caring manner, and keeping me on track. I want to thank many of my friends and colleagues, some in and out of the program. Thank you, Dr. Emily Brown, Dr. Jim Furstenberg, Dr. Molly Cooper, Dr. Jonathan Adkins, Tommy Pollack, Amy Antonucci, and a host of others. I appreciate your support. Greatness occurs when others contribute to academic support and mental enrichment.

Second, I wish to thank my professors and dissertation committee members for providing invaluable feedback during my proposal defense. Additionally, I am thankful for my dissertation chair, Dr. Ling Wang, and her time devoted to reviewing my documents over the years. Dr. Wang's feedback and openness not only improved the final report but often positioned my thought process to include additional and pertinent content, all relevant to the research.

Third, I wish to thank my editor Elizabeth Conard with Editide. Elizabeth spent many hours editing and suggesting changes that were conducive to improving this document. Elizabeth's background as an English major played an instrumental role in moving the final report to completion.

Fourth, I want to thank Carol Akins for putting up many times with me saying, "I need to work on my research." I know I couldn't travel and needed quiet time for research; not everyone understands this.

Lastly, Nova Southeastern University staff did an outstanding job during the COVID-19 pandemic by ensuring communications remained updated and kept online access to the libraries and literature. There are always people in the shadows who support our endeavors that need to be recognized.

Table of Contents

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Background

Insider threats pose significant security risks within organizations; this specific threat is nothing new. The Intelligence and National Security Alliance (2015) defines insider threats as follows:

> The threat presented by a person who has, or once had, authorized access to information, facilities, networks, people, or resources; and who wittingly, or unwittingly, commits: acts in contravention of law or policy that resulted in, or might result in, harm through the loss or degradation of government or company information, resources, or capabilities; or destructive acts, to include physical harm to others in the workplace. (para. 3)

Organizations place focus on monitoring solutions and appear to operate reactively. Proactive approaches to security may be more beneficial than reactive approaches (Hunker & Probst, 2011). Statistics have captured reported instances of insider threats; however, some cases are not reported due to ethical laws and sanctions issued by the government (Oladimeji, Ayo, & Adewumi, 2019).

To this end, this dissertation study does not represent an end-all solution to insider threat mitigation. Instead, it addressed why past attempts to mitigate insider threats have failed and provided an alternative approach to alleviate the problem by leveraging data from social platforms. Organizational leaders could ascertain whether an individual might

be considered a security threat at the earliest onset and before employment within this context.

Problem Statement

Insider threats are a problem. Powers' (2017) brief literature review documented a case of an insider threat in the financial sector in 1792. Other forms of insider threats dated back to 41 A.D.

This dissertation study's review of relevant literature revealed various avenues and approaches to standard practices to address insider threat mitigation efforts; however, the literature fell short of effectively addressing the problem. Organizations have invested heavily in deterrence monitoring tools to observe employees' activities, such as computer access, Internet browsing, and e-mail communications (Alahmadi, Legg, & Nurse, 2015), yet requires monitoring resources, and can become an ineffective strategy. According to Loffi and Wallace (2014), employee monitoring for insider threat activities should be used cautiously with concerns for lapses in enthusiasm and suspicion of leadership.

To further combat the insider threats, Cole (2015) suggested that organizations implement administrative policies, procedures, Internet audits, workforce monitoring, whistleblower incentives, and put strategies in place for data loss prevention. Individuals who cause insider threats are aware of the policies, procedures, and technology used in their organizations and are often also aware of the organization's vulnerabilities (Cappelli, Moore, Trzeciak, & Shimeall, 2009). Auditing can also impact policy enforcement. Auditing often results in redundant, misleading, and missing data; even worse, audit trails typically lack time correlation (Hunker & Probst, 2011).

Effective information security cannot be delivered only by perfecting the effectiveness of technical controls (Ismail & Yusof, 2018) and goes beyond technological aspects. Attempts to mitigate an insider threat that is solely based on technology could fail and operates during or after a threat. Methodologies of the past had weaknesses, exploitable flaws, and demonstrate a need to foster new approaches.

This dissertation study leveraged an additional source of data to supplement existing mitigation efforts. As outlined in the research process, the IT artifact creation (see Figure 1) included information collected through fraudulent legal cases. In these instances, the social components from fraud was leveraged to provide a better understanding of the cohesion within social media data (with permission of the author).

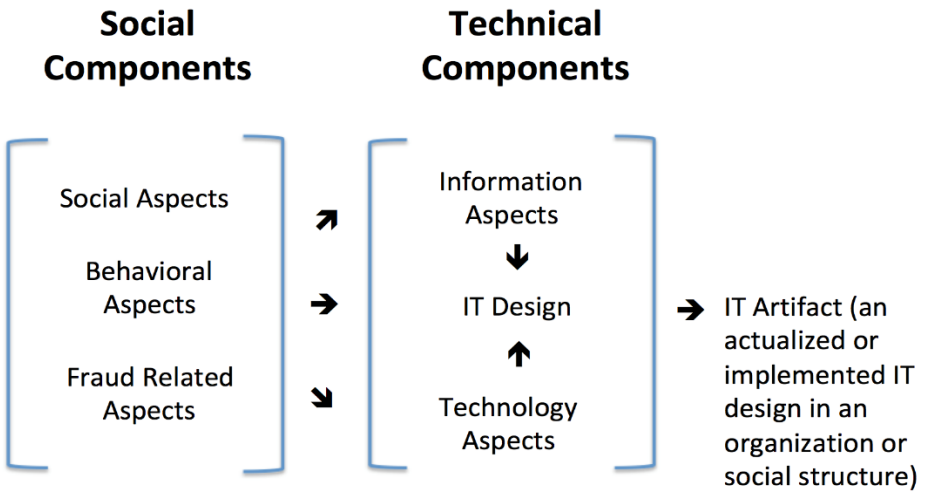


Figure 1. Modified IT artifact. Reprinted from "Considering the social impacts of artefacts in information systems design science research" by G. De Leoz and S. Petter, 2018, *European Journal of Information Systems*, *27*(2), 154–170. Copyright 2018 by Taylor and Francis. Adapted with permission.

Many types of insider threat classifications must be considered. Shaw, Ruby, and Post (1998) pointed out vulnerabilities associated with introversion, computer dependency, social and personal frustrations, ethics, entitlement, lack of empathy, and

reduced loyalty. The classifications can become exhaustive; therefore, the focus of this study was related to generalized fraudulent activities through court transcripts.

Court transcripts with cases involving forms of fraud, illuminated schemes, and tactics used in carrying out crimes and shows relevance to this study's interpretation with behavioral theories. For example, a corrupt security executive's scheme included giving himself his regular paycheck, then later would write a second paycheck, a forged check (Glackin & Bible, 2019). The case blatantly lists the defendant's fraudulent behaviors through multiple counts of felony theft and forgery. In this instance, information on fraud is also illustrated through court documentation and appears to be a great source for supportive information.

While an initial assessment into insider threats is understood, ways with dealing with the threat appears to remain a serious problem worth solving. Data collections from various sources can hold the key to detecting the threat from much earlier in the mitigation cycle. The integration of innovative data collections from the courts, analysis of social media data, and applying the behavioral theories can provide a clear picture to represent possible insider threat characteristics.

One could postulate that data extracted specifically from prosecuted cases involving various forms of fraud can show relevancy to different behavioral theories and reveal correlations to social media posts that exhibit specific negative sentiment within comments, messages, or user-posted content. It is not guaranteed that one will find consistent matches of sentiments; however, it is possible to use machine learning and arrive at a probable outcome.

After reviewing what has already been done, the problem remains. Extending research by Park, You, and Lee (2018), appears to offer new light when integrating negative sentiments from fraudulent data and into the context with social media. The added value obtained from leveraging legal documents provides useful information in identifying threats. For instance, data supporting the research was achieved through leveraging top website outlets, such as sites reviewed by PracticaleCommerce (2017), and included access to twenty-thousand Twitter tweets. Additionally, court transcripts came from several jurisdictions around the country and accounted for approximately 138 pages from legal documents.

This dissertation study sought two types of data. First, accessing social input tweets from Twitter provided insight into social contexts. Insiders have shared common characteristics, and the extraction of these attributes through social media is feasible (Gritzalis, Stavrou, Kandias, & Stergiopoulos, 2014). Tweets have varying types of content. The following are starting points for capturing content:

- Date
- Time
- Source
- Favorites
- Retweets
- Replies
- Quotes
- Language
- Tweet types

- Text
- Quote
- Country
- Place
- Latitude
- Longitude

Second, the need to access court transcripts included a preliminary search for a suitable tool and led to the Public Access to Courts Electronic Records (PACER) website. PACER provides case information from the eastern District of Virginia, including Albemarle, Fairfax, Loudoun, and Prince William Counties (USCourts.gov., 2019). Moreover, the Loudoun County General District Court (LCGDC) offered public computer access and made cases available. S. Shifflett from the LCGDC office stated that the LCGDC database allows researchers the ability to view most filings and then determine any required documents (personal communications, August 23, 2019).

In addition, questions from the prosecuting counsel and the defendant's replies to counsel are intentionally omitted. This dissertation study needed a preview of case information to understand what information was available, which required an in-person visit to LCGDC, and sought the following preliminary types of information.

- Prosecutor's questions
- Defendant's responses to questions from prosecuting attorneys
- Plea agreement(s)
- Arraignment
- Complaint

- Filing
- Probation
- Readiness
- Sentencing

Social media inputs and court transcripts provide insight into an individual's character. Furthermore, social network data is beneficial for detecting threats at a much earlier stage (Kauh et al., 2017); thus, social network data became a factor leveraged in this research. Park et al. (2018) found practicality in social media analysis. Albeit, this study used contrasting data to find negative emotions in support of discovering insider threats, with emphasis placed upon preemployment vetting.

While the preceding works by Kauh et al.'s (2017) offered promise, attempts to prevent an insider threat appeared to leave room for improvement. The research intent extends similar works by Park, You, and Lee (2018) in broadening social media sentiment analysis to include additional data sources that show relevancy to the Routine Activity Theory (RAT). More importantly, fraudulent case data contained lexicons with correlations to various forms of fraud are sought after within social inputs. Emphasis was placed on early detection, as it is nearly impossible to stop the insider right before the incident; hence the best solution is to prevent the threats from occurring through early detection (Soh, Yu, Narayanan, Duraisamy, & Chen, 2019).

This dissertation study used a revised approach and warranted an in-depth screening of individuals' online activities to determine each individual's level of trustworthiness. The idea was to provide another tool to be used in a rounded approach to mitigate an organizational threat during the employment vetting process.

Dissertation Goal

The study's goal is to explore the various theories (Table 1), show which relates to fraudulent cases, examine social media data streams, and examine correlations through machine learning of sentiment analysis. Supportive behavioral theories provide another mechanism to supplement a well-rounded approach in predicting an insider threat. Just as court transcripts offered a wealth of information with specific fraud traits, social media content is another instrumental source of data used to assess an individual's character. According to Kandias, Stavrou, Bozovic, and Gritzalis (2013), when employees exhibit antisocial and negative views of law enforcement and those in authority, they become more likely to act against an organization; traits are becoming increasingly important when attempting to identify an insider threat.

Inputs from social media can fill a gap with applicants not being forthcoming with certain information and can potentially decrease their realistic or imagined chances of employment (Jeske, Lippke, & Shultz, 2019). Analogous to this study are the efforts conducted for the U.S. Department of Homeland Security (DHS). The agency sought a contract to build an "extreme" vetting system that would analyze social media posts (Duarte, Llanso, & Loup, 2018) and demonstrated a desire to do more with mitigation.

Table 1
Referenced Theories

Theory	Reference
General deterrence theory	GDT
Protection motivation theory	PMT
Routine activity theory	RAT
Social bond theory	SBT
Theory of planned behavior	TRB
Theory of reasoned action	TRA

Leveraging social media content is nothing new. Eighty-six percent of employers screen prospective employees' social media content (Berski, 2016), whereas others validate information provided on a résumé (Carpenter, 2017). Furthermore, social media checks are becoming increasingly important and should be pursued (Kühn & Nieman, 2017).

This study focused on the earliest phase of the preemployment vetting process, beyond a manual intervention of viewing profiles and résumés. The model for this research led to a theoretical framework that included information from social inputs and publicly available court transcripts from prosecuted fraud cases.

The usage of court data, applying theories addressing behavioral norms, and social media inputs produced an attainable concept to promote and support a mitigation effort. Data supporting the research were available through the court systems and through diverse website outlets, such as sites reviewed by Mehra (2017) and access to thousands of users in social media.

Park, You, and Lee (2018) shared a similar purpose with research leveraging social inputs to focus on individual behaviors when examined through social content. Moreover, changes in behavior or mindset and attitude are often displayed either before or during the insider act being committed (Bell, Rogers, & Pearce, 2019). Analyzing social inputs presented itself as an effective use of data, provided in-depth insight into improving the vetting process, and ultimately allows organizations another tool to help prevent the selection of anyone likely to position themselves as an insider threat.

Research Questions

This dissertation study formulated the research questions to address whether or not the usability of behavioral theories, fraudulent court transcripts, and social media inputs could be leveraged as tools used within the preemployment hiring process. The following research questions directly correlated to the design science research's (DSR) artifact process, which involves categorizing each question into either a social component, a technical component, or both.

RQ1: Is there sufficient literature on insider threat mitigation strategies? (Technical component).

RQ2: Is there relevance in behavioral theories, court transcripts from fraudulent cases, and social inputs that can solve the problem with the research? (Social component).

RQ3: What behavioral theories are most applicable to the research? (Social component).

RQ4: Can IT artifacts be created from the information obtained in behavioral theories, from court transcripts of fraudulent cases, and social inputs? (Technical component).

RQ5: Will each IT artifact yield favorable outcomes through lab experiments and contribute to the goal of the study? (Technical component)

Relevance and Significance

The markets most significantly impacted by insider threats include U.S. banking, finance, information technology (IT), healthcare, government, and commercial facilities (Williams, Levi, Burnap, & Gundur, 2018). Other likely outcomes beyond a monetary

loss include serious harm to the organizations' confidentiality or integrity (Krull, 2016), embarrassment, legal fines, and loss of competitive advantage (Williams et al., 2018), to exposure of customer data, trade secrets, or even leaking classified information.

Two highly documented cases in the intelligence community include Chelsea Manning, who was convicted of stealing and disseminating 750,000 pages of documents and videos to WikiLeaks (Jarrett & Borger, 2017), and Edward Snowden, the U.S. National Security Agency contractor who leaked classified information in 2013 (Kühn & Nieman, 2017). In both instances, leaks of classified information presented a significant threat to national security.

There is a belief the problem of insider threats may exist due to a relaxed vetting of individuals—such as current or former employees, contractors, or business partners (Park, Lim, Kwon, & Choi, 2017)—who act outside the trust expectations that others in the organization set (Costa, Albrethsen, & Collins, 2016). The sense of slackened onboarding can supply a theoretical significance with the psychological reasoning to answer questions centered around human behavior; thus, onboarding is a key aspect in threat mitigations.

Moreover, the RAT states that criminal acts require convergence in space and time of likely offenders, suitable targets, and the absence of capable guardians against crime (Cohen & Felson, 1979). For example, Williams et al. (2018) provided first-hand evidence of the routine activities and guardianship that play a key factor in the likelihood of insider threats. Many theories, such as the general deterrence theory and the theory of planned behavior (TPB), provide an in-depth understanding of an individual's intentions

with information security compliance practices established by organizations (Flowerday & Tuyikeze, 2016).

Integrating applicable behavioral theories to portions of this research is primarily done through the researcher's interpretations and observations extracted from fraudulent case data and court reporting. The observations were taken from supporting documents are taken through the researcher's view and systematically sorting through the data to find common themes (Creswell & Miller, 2000), relevant to the research conclusion, and becomes a portion of the ending report. The preceding provided another aspect that carried significant weight in understanding why insider threats occur and examined the context in which the threats operate and shows relevancy to current-day organizational threats.

The study results addressed insider threats at an earlier stage, making it feasible for organizational leaders to mitigate attacks through the analysis of social inputs and indicate the likelihood of an individual being a potential insider threat (Alahmadi et al., 2015). Advanced recognition of a threat can become a tool that organizational leaders can add to their comprehensive personnel vetting approach. Moreover, an effective defense against insider threats is more of a result from a multipronged approach (Catrantzos, 2018), and not based on a single methodology.

Additionally, the research findings contribute to the body of knowledge through additional analysis of social media content, which demonstrates a conducive direction for detecting insider threats from an earlier position, and focused on how supplemental data sources can add value to future research.

Barriers and Issues

Several identified barriers and issues are attributed to data collection within the context of gathering court transcripts. Although the initial gathering of court transcripts came through the PACER system, knowing case numbers became a lengthy process. According to USCourts.gov (2019), there are approximately 10,000,000 criminal cases and filtering to a specific region can become a daunting task.

Due to the volume of available data, restrictions needed to be implemented during the data collection phase, requires paralegal support, and support from the local commonwealth attorneys. If paralegal support and support from the local commonwealth attorneys remains an obstacle, legal counsel suggests not working with court transcripts due to the volume of information; instead, researchers should use case notes or briefs of cases from court reporters to alleviate the previous concerns (B. Gilliam, personal communications, September 15, 2019). Court reporters must get permission to release transcripts to anyone who is not a party or participant in the case. However, court reporters need to get permission to release transcripts to anyone that is not a party or participant in the case; it is not to say they cannot or would not (D. Linton, personal communications, November 11, 2019) be available. According to Jaafari & Lewis (2019), fourteen states have replaced court reporters with technology capable of capturing audio and video.

After the court transcripts became available, additional speech to text conversion software such as Bear File Converter, or a similar tool, played a crucial role in converting the material. In addition, PDF files later required using iSkySoft to convert into text, and later discussed with the lab experiments. The tools identified in the study required

sufficient time to understand, which was best fitted for the tasks and capable of performing sentiment analysis from various data streams. Some tools require conversions of portable document format (PDF) documents into files that are compatible with the other tools used within a study; thus, PDF-formatted documents and documents including bitmap images create challenges during data extraction (Staar, Dolfi, Auer, & Bekas, 2018). Nevertheless, issues with document conversions are overcome through a rigorous trial-and-error approach. The absence of electronic documents required converting hard copies by scanning into a digital format.

Sufficient time and resources are needed and should be dedicated to the preparation, test sampling, cleansing imbalanced data sets, and working to validate the efficiency of planned techniques for dealing with the class imbalance problem in big data sets (Krawczyk, 2016; Patil & Sonavane, 2017). The significance with imbalanced data is voluminous, especially as social media data is growing with ever-increasing needs to analyze large amounts of data to get useful insights (Kamburugamuve, Wickramasinghe, Ekanayake, & Fox, 2018).

Furthermore, this research must account for the correction of imbalanced data retrieved from social input from Twitter within the lab experiment portion of the DSR methodology. In this instance, this dissertation study centered the data update around limiting the number of records being processed, which was a constraint among some of the study's tools.

Lastly, due to the Coronavirus (COVID-19), visiting local jurisdictions often required communicating in advance to verify hours of operation, following specific protocols when entering office buildings, such as maintaining social distancing, and

wearing personal protective equipment. Other means of communicating included electronic mail and often required the need to be mindful of the different time zones, especially when seeking prompt responses.

Assumptions, Limitations, and Delimitations

The assumption with the study included the following. Collections of Internet data would not become an arduous process. However, collecting data from public Internet sites such as Twitter relies on general access from third-party tools, and obtaining access presented itself with unique requirements. This dissertation study's preliminary efforts included uncovering the tools or vendors used in the study that currently had APIs to support data retrieval. Twitter required a developer's account and the company must vet the application; unfortunately, time ran out while waiting for this process to complete. Due to this obstacle, this dissertation study leveraged services through another vendor—Vicinitas (2020)—to provide Twitter tweets.

An additional assumption was derived from the unknown amount of groundwork that was needed during the gathering of documents. Document conversions can be delayed until court transcripts are reviewed in-person. For instance, documents might only be available in a printed format and not in the computer-readable format of PDF. The required time to convert was an unknown factor, and the level of difficulty was outlined in the barriers and issues' section of this chapter.

Another area classified as a limitation came from the research using data written in English and applied to sentiments, Twitter tweets, and court data. More than 3,600 interpreters are registered in the judiciary's National Court Interpreter Database; these interpreters cover 180 languages routinely used by the courts (USCourts.gov, 2017). This

dissertation study did not want to add another layer of complexity involved in translations within the court system. Facebook alone supports 111 diverse languages to its user base, with Twitter promoting at least 33 languages (Fick & Dave, 2019). It is possible to extend these works towards other languages; however, this was out of this study's scope.

An initial assessment of delimitations is established by a finite number of artifacts used in the study. Ideally, purposive sampling is used because it is a sample chosen "on purpose" because those sampled meet specific criteria (Terrell, 2015) and correlated with the number of fraud cases used in the research. Conversely, it was possible to extend the delimitations into additional classification types of court cases, and present each artifact with varying degrees of uniqueness that goes beyond fraud.

Definition of Terms

Big data - Collection of a very large amount of data (possibly in terabytes or even bigger) being generated by numerous users all around the world via different instruments and technologies (such as the web). It is difficult for technologies that process small amounts of data to handle, process, analyze, capture, and visualize big data (Patil, Kamdar, & Khatri, 2014).

Discriminative Multinomial Naïve Bayes – According to Panda (2018), Discriminative Multinomial Naïve Bayes is based on a Naïve Bayes variant with emphasis towards characteristics of discriminative learning within text classifications.

Lexicon - A lexicon is a book containing an alphabetical arrangement of the words in a language and their definitions (Merriam-Webster, 2020).

Naïve Bayes - Naïve Bayes is a probabilistic based supervised learning algorithm that uses Bayes rule together with a strong assumption that the attributes are conditionally independent, given the class (Webb, 2010).

Random forests - Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001).

Support Vector Machine - An SVM is a kind of large-margin classifier: it is a vector space based machine learning method where the goal is to find a decision boundary between two classes that is maximally far from any point in the training data (Manning, Schütze, & Raghavan, 2008).

Valence Aware Dictionary and Sentiment Reasoner - Valence Aware Dictionary and Sentiment Reasoner (VADER) is a parsimonious, rule-based model for sentiment analysis of social media text (Hutto & Gilbert, 2014). Within the context of this research, VADER's gold standard list of lexicons provided scoring for words focused on negative content directly attributed to tweet sentiment.

List of Acronyms

ARFF	Attribute-Relation File Format
AUP	Acceptable Use Policy
CSV	Comma-Separated Values
DMNB	Discriminative Multinomial Naïve Bayes
DSR	Design Science Research
EPA	Electronic Public Access
FPR	False Positive Rate

HR	Human Resources
IT	Information Technology
LCGDC	Loudoun County General District Court
OCR	Optical Character Recognition
PACER	Public Access to Court Electronic Records
PDF	Portable Document Format
RAT	Routine Activity Theory
SIEM	Security Information and Event Management
SVM	Support Vector Machine
TPB	Theory of Planned Behavior
TPR	True Positive Rate
TRA	Theory of Reasoned Action
WEKA	Waikato Environment for Knowledge Analysis
VADER	Valence Aware Dictionary for Sentiment Reasoning

Summary

In summary, Chapter 1 discussed the background of insider threats, the industries significantly impacted, and proposed a way forward based on past research. Specifically, proposing organizations move from a reactive posture to a proactive stance by examining sentiment analysis of social media inputs and court case information relating to fraud. The end goal of this study included determining specific threat levels; this process could become another tool in an integrated approach to mitigating an insider threat during a vetting process. The lengthy and tenuous history with insider threats appears to focus on degrees of monitoring and policy enforcement. For instance, monitoring and placing

mechanisms that must be observed appears to postpone the inevitable, thus delaying the threat. Therefore, prudent security involves more than perfecting the effectiveness of technical controls (Ismail & Yusof, 2018).

Existing works can be extended by integrating sentiment analysis with other forms of data and adding social attributes to create value in the vetting of individuals through a more aggressive and proactive measure. This dissertation study had three goals for conducting this study: (a) to analyze data from new sources to include court cases centered around fraudulent activities that is applicable within social media Twitter tweets, (b) to provide another vehicle in vetting individuals that go beyond natural language processing (NLP) shortcomings with filtering social media posts for dangerous content, and (c) to deepen analysis through the merging of relevant data attributed with varying degrees of fraud. The bridging of case notes and transcripts into the social realm provided a profound advancement with mitigation efforts with insider threats.

Chapter 2

Review of Literature

The literature review in this section covers content that directly supports the research with a close look into monitoring shortcomings, problematic rulemaking, policy enforcement, and weak vetting practices with onboarding employees. For instance, a review of self-disclosures is examined as part of the applicants' hiring and vetting process. Researchers who conducted ancillary literature reviews considered monitoring techniques to mitigate an insider threat, but monitoring can become more of a reactive and labor-intensive means of taking corrective measures. The weaknesses from previous insider mitigation approaches places emphasis to extend similar works to Park et al. (2018). This dissertation study's research is analogous to works from Gritzalis et al. (2014), who combined social media data to detect both technical threats and threats associated with theory-based behavioral changes.

This dissertation study selected reviews that addressed existing insider threat mitigation efforts. Insider threats are defined as a threat typically attributed to legitimate users who maliciously leverage their system privileges and familiarity and proximity to their computational environment to compromise valuable information or inflict damages (Chinchani, Iyer, Ngo, & Upadhyaya, 2005).

This dissertation study examined literature directly associated with mitigation strategies that provided a more precise understanding of the problem. From a foundational point-of-view, Cappelli et al. (2009) suggested that measures that are in practice today should include best practices in understanding the threat is organizational-wide.

As illustrated within DSR (see Figure 2), the research process incorporated theories to paint the picture of the insider threat. Leonard, Cronan, and Kreie (2004) hypothesized that behavioral intention is influenced by an individual's attitude, which in turn is influenced by consequences of action and the environment, obligation, and personal characteristics. Theory-based contributions within the context of this study augmented other technical and social components of the research.

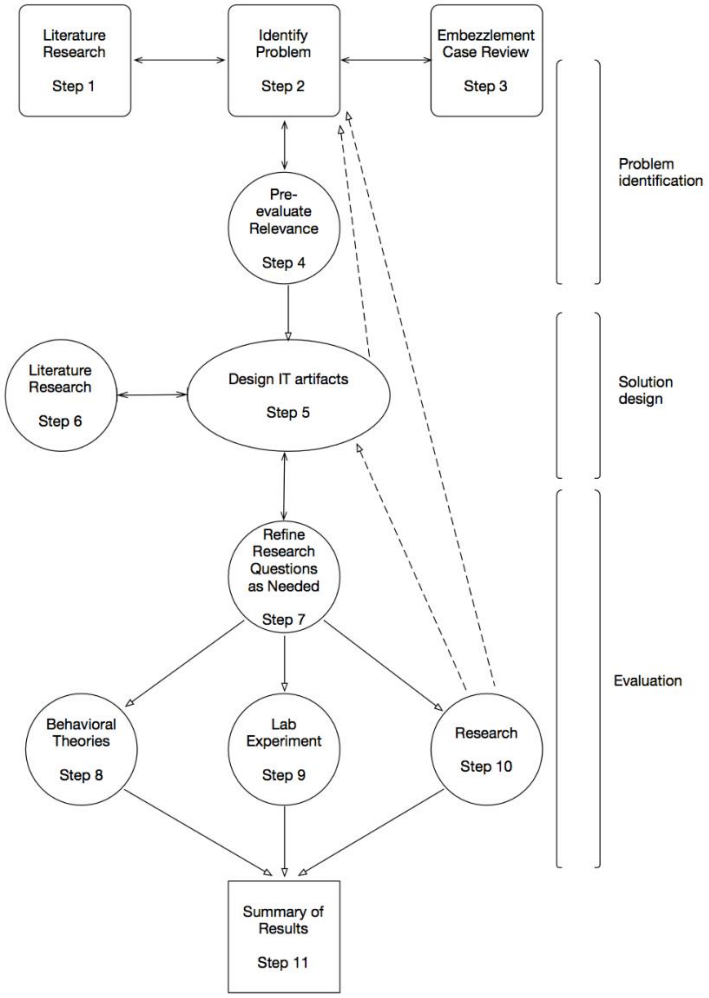


Figure 2. Research process. Reprinted from "Outline of a design science research process," by P. Offerman, O. Levina, M. Schönherr, and U. Bub, 2009, In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, 1–11. Adapted with permission.

It is crucial to extend relevant works, capitalize on social media sentiment analysis (with permission of the author), couple fraudulent case data within the legal system, and provide a newer mitigation strategy for insider threats. The following sections focus on what is being done in respect with monitoring, profiling, rulemaking, policy enforcement, and employee vetting practices, all with their own weaknesses.

Monitoring and Profiling

Monitoring accounts comprise a significant portion of mitigation strategies and includes internal network monitoring, external monitoring, and employee monitoring (Cole, 2015). However, the price tag associated with continuous monitoring is not apparent. Additionally, it is not practical to institute daily monitoring (Cappelli et al., 2009) without being mindful of required resources.

Other researchers focused on user profiling with the goal to mitigate insider threats by profiling user activities such as capturing keystrokes, monitoring web browser activity, files accessed, removable media, and USB activity. Commands such as Change Directory (CD), Print Work directory (PWD), List (LS), Copy (CP), and Remove (RM) are construed as activities that might be labeled as potentially malevolent (Liu, De Vel, Han, Zhang, & Xiang, 2018). Often, logging user activities with deep analysis has offered insight into anomalies to address new observations, times of observations, and frequency of observations. All known cases exhibit a change in user behavior (Legg, 2015).

Similarly, Shaw (2006) conducted observation-based profiling and discovered through employee profiling, found risk indicators with management labeling workers as difficult, and often disgruntled with other employees. A higher percent of those being profiled attracted close observations from those in supervisory positions just before security incidents. In these instances, identifying the insider threat is flawed when factoring in time to process; this became a common theme in Shaw's research. Mitrou, Kandias, Stavrou, and Gritzalis (2014) stated the following in regard to monitoring social media:

> Online social media profiles, blogs, tweets, and online fora are increasingly monitored by employers searching for information that may provide insight on

> employees and prospective hires. Taking into consideration the exponentially growing participation in online social networking sites and social media, it is not surprising that employers are searching for unique information about applicants and employees not found with other selection methods. (p. 9)

Monitoring solutions can become costly; however, monitoring solutions creates a barrier and deterrent for many organizations that need to implement an insider threat program (Spooner, Silowash, Costa, & Albrethsen, 2018). Organizations may not benefit by investing significant amounts of time watching for a collection of events through log analysis or through implementing intrusion detection systems. Kauh et al. (2017) developed an insider threat model that was capable of inspecting threats within network packets, with the long-term research goal of detecting insider threats within a network.

From a network perspective, thwarting an insider threat leaves more to be desired when operating in real-time. Operators must review system attacks without a reference to anything that is already on file, often becoming ineffective due to changes in the environment (Benferhat, Boudjelida, Tabia, & Drias, 2013), and can lead to undesirable results. Similarly, data taken from analyzing security information and event management (SIEM) content yields problems with parsing data from security logs, making it difficult to design a detection policy for security threats (Lee & Huh, 2019).

Spooner et al.'s (2018) SIEM study demonstrated capabilities to help mitigate insider threats through anomaly detection and provide evidence to support legal actions. However, improperly implemented SIEM benefits result in cumbersome and undesired tendencies, leading to system issues with log aggregation.

SIEM solutions often come tethered with many problems, such as inconveniences with log management, reporting, real-time monitoring, integration and deployment, product quality, and stability (Splunk, 2020). In addition, organizational staff require

training, product familiarity, effort, and expertise to implement successfully. An exhaustive suite of tools supports insider threat mitigations; however, some out-of-the-box solutions are somewhat useless and require a significant understanding of the distinct intellectual property generated across organizational components (Spooner et al., 2018).

False-positive alerts should be addressed for SIEM-related technologies to work; otherwise, this type of technology could overwhelm analysts, making tasks ineffective and inefficient. In contrast, precondition or post condition may be missed due to false negatives (Hubballi, & Suryanarayanan, 2014), which is the absence of alerts in the presence of attacks (Kenazag, Tayeb, Mahdi, & Aiash, 2016). Moreover, a successful implementation requires competent security personnel, with focused efforts to correlate rules to drive down response times and work towards minimizing false-positive alerts (Vilendečić, Dejanović, & Ćurić, 2017).

Incorrectly implemented technology places the burden on employees. It is unrealistic to expect an individual to pore over voluminous log files on a daily-basis (Spooner et al., 2018). The laborious effort to review logs contributes to delays in identifying threats, which is not a viable solution in preventing an insider threat. While all aspects to monitoring appear as practical approaches, the process of identifying the threat is not always real-time; thus, preventive action cannot be taken at the right time (Ambre & Shekokar, 2015), leaving the insider threat a lingering concern.

The literature demonstrates a wide latitude of methodologies, including surveys, business cases, machine learning models, sentiment analysis, supervised learning, unsupervised learning, theory-based, and a host of others. Several methodologies stood out in unexpected ways. For instance, Bell et al. (2019) used a methodology that involves

using a survey to provide an in-depth understanding of behavioral indicators. Bell et al. found the changes in behavior, mindset, and attitude are often displayed either prior to or during the insider act being committed. In contrast, Williams et al. (2018) conducted another empirical study and fostered a theory within their survey that correlated with the range of crime issues connected to the RAT. Williams et al. posited that theories can be applied to insider cyber victimization.

The issues that result from monitoring practices generate concerns over ethics and privacy. Acceptable use policies (AUP) are among the most common company policies that outline how employees can use company systems and what employees can expect in regard to privacy (Yerby, 2013); however, these policies do not appear to mitigate insider threat activities. Providing an acceptable use policy is worthless if the employees do not become aware of them (Alshboul & Streff, 2017). Also, without a successful implementation, does not change users' attitudes and behaviors, and makes no impact on mitigating insider threats (Gallagher, McMenemy, & Poulter, 2015).

Although technology cannot solely guarantee a secure environment for information, the human aspects of information security should be taken into consideration (Safa, Von Solms, & Furnell, 2016). The importance of addressing the insider threat nontechnical component, moved this research towards leveraging social inputs to drive the significance through sentiment analysis further.

Rulemaking and Policies

Rulemaking can impact the human and technological aspects of monitoring for insider threats (Spooner et al., 2018). Linkov, Poinsatte-Jones, Trump, Ganin, and Kepner (2019) postulated both over-regulation and under-regulation can be exploited by the

insider threat. An optimal solution to rulemaking and policy enforcement could require the knowledge to understand the number of rules, instead of measuring the context of rules. Rules include exceptions due to higher authority principles (Antoniou, Billington, & Maher, 1999); these exceptions often override older regulations and could present an unclear direction to employees who rely on voluntary compliance and cooperation (Pelton, 2017).

Other forms of rulemaking are presented as organizational policies and outlined in several case studies. Bauer, Bernroider, and Chudzikowski (2017) discovered that many organizations have policies specifically addressing internal threats; nevertheless, individuals intentionally act noncompliant. Supplemental measures become ineffective when responsible personnel violate or override the policies and procedures, irrespective of whether this is caused by carelessness, poor knowledge, or clear intention to act dishonestly (Nawawi & Salin, 2018).

Comparatively, Cram, Proudfoot, and D'Arcy (2017) found that organizations exhibited a lack of continuity with security policies and demonstrated a lack of cohesion with its employees. The most prevalent relationship within Cram et al.'s framework is the relationship between enforcement difficulties, excessively complex policies, inadequate resourcing, and failure to customize policies. Given the voluminous of inefficiencies in successful policy implementations, the rulemaking variant in a deterrence does not appear to be a workable solution in preventing the insider threat.

Employment Vetting

The prevention of an attack is just as important as other components to supplement existing practices; thus, prior researchers have examined various

methodologies centered around employment vetting. Current vetting mechanisms are slow and less capable of catching new threats (Chen et al., 2015); yet, companies continue to rely on these ineffective practices. The vetting of applicants strengthens the collective efforts in the overall vetting; however, flaws remain due to the over-reliance on information obtained from the employee (Kühn & Nieman, 2017).

Edward Snowden is a prominent example of flawed employment vetting. Snowden was a contractor who leaked classified information in 2013 and sailed through multiple security vetting interventions (Kühn & Nieman, 2017). Simpson and Foltz (2017) discussed their concern for the lapse of vetting activities between vetting cycles and a contributing factor. These weaknesses result in unnoticed recognitions in detecting the trustworthiness of individuals. The vetting process's flaws include the lack of vetting the vetting officials, and the tendency for vetting officials to demonstrate a level of bias, interject, and intertwine personal experiences within the formal hiring process, thus altering vetting outcomes (Lomas, 2019).

Conversely, employees subjected to polygraphs do not always have questions best suited to the position for hire; therefore, security vetting investigators can hinder the detection of misconduct (Kühn & Nieman, 2017). Jeske et al. (2019) found that voluntary disclosures during preemployment indicate a prospective applicants' willingness to trust, privacy concerns, and perceived a vulnerability associated with the use of information about applicants. These indicators may be important predictors of self-disclosure involved in information sharing. Even so, the decision to disclose is sometimes forced upon the employee with little warning, potentially after the hire, and does not appear to solidify sound practices (Hielscher & Waghorn, 2015).

Conventional vetting practices that rely on honesty are flawed because employers are unable to verify the information or determine which information they should not consider in their decision-making process (Jeske & Holland, 2019). It is possible for those being interviewed to express some level of undesired personality traits (Roulin & Bourdage, 2017). Some traits are classified as misleading and considered a potential threat to businesses, which depicts a dark picture of organizations' ability to deal with the threat represented by applicants' use of deceptive impression management tactics (Roulin, 2016).

Maasberg, Warren, and Beebe (2015) examined insider threats based on personality trait profiling during the hiring process. Maasberg et al. aimed to build propositions of personalities and factored in negative attitudes, malicious intent, triggers, motives, capabilities, and opportunities centered around security weaknesses. Moreover, it would appear unmasking the preceding during postemployment, offers even less of a solution in mitigating an inside threat.

Organizational leaders must focus on the prevention of insider threats in the earliest stage. BaMaung, McIlhatton, MacDonald, and Beattie (2018) suggested using a comprehensive and intrusive approach at the earliest onset.

Summary

Previous studies exhibit some level of researched solutions to mitigating insider threats; however, all strategies appeared to operate with constant monitoring, act in a reactive state, or rely on employees' honesty. The crux to this research places the attention on the human resources (HR) preemployment phase. Preventing potential threats should be HR's central issue of concern (Fischbacher-Smith, 2015), and the

practice of social media vetting can afford the employer access to information about the candidate that they might not otherwise find (Delarosa, 2015).

From a prevention point of view, some tools and methodologies examined from prior research leave more to be desired. Past studies have considered plans for continuous monitoring. Employee monitoring was relatively ineffective for some (Wallace & Loffi, 2014) because this type of monitoring required security expert knowledge of SIEM rule creations. Other studies indicated that employees who do not understand the correct balance of security rules within an organization, will not follow the company's guidance, and become the precursor to insider threats. Past mitigation strategies have appeared to fall short of genuinely preventing the insider from entering an organization.

With the advent of social media, there appeared to be a value when analyzing content for negative sentiments. Gritzalis et al. (2014) postulated that online content could reveal those who have demonstrating traits of an insider threat. Various researched approaches from previous studies indicate that sentiment analysis and machine learning are likenesses to the research from works by Park et al. (2018). However, the main difference between that study and this is based upon identifying negative sentiments that directly correlate to fraudulent judicial data extracted from the legal system. Previous studies' inclusion of social media data in efforts to mitigate insider threats provided the foundation for the present study. Such information contributed to the relevancy in moving away from reactive demeanors to proactive measures.

Chapter 3

Methodology

Overview

This dissertation study used a design based on DSR to explicitly create IT artifacts that had unique attributes to promote preemployment security vetting against insider threats. The artifacts included data extracted from prosecuted fraud cases from the court system and applied sentiment analysis by combining said data to social inputs, primarily tweets. Supportive to data collecting came two different sources of data to (a) build negative lexicons from the court documents, (b) correlate the same negative lexicons with social media tweets, (c) perform sentiment analysis within the selected tweets, and (d) discover behavior theories applicable within the fraudulent court data; discussed in subsequent sections.

The present study's methodology was similar to research conducted by Zaib, Asif, and Arooj (2019). Zaib et al. focused on word and sentence tokenizing and provided a partially based model on implementing tidytext. Zaib et al. used tidytext to collect the most negative comments and compare word and sentence analysis to tune their approach. Conversely, Silge and Robinson (2017) found that tidytext included functions and data sets capable of text conversions that could be integrated with existing text-based mining packages.

Hutto and Gilbert (2014) evaluated thousands of unique lexicons and concluded that VADER was a top contender in regard to speed and capability. It is possible to leverage existing VADER lexicons through some form of a modified hybrid approach. For instance, one can update sentiment lexicons through compiling Waikato Environment for Knowledge Analysis' (WEKA) word lists in a manual process. The manual updating of negative lexicons requires updating the polarity scores, editing positive and negative indicators, the actual lexicons (Bonta & Janardhan, 2010; see Appendix A), and through the usage of custom created scripts.

Research Methods Employed

The research approach was based on DSR (Offermann et al., 2009). This dissertation study based the IT artifact's process design on works by De Leoz and Petter (2018); this process design is encapsulated within the DSR.

Step 1 in the initial research process involved conducting the literature review on insider threat mitigation strategies. The conducting of literature review had two goals: This dissertation study's goal was to review what has already been done and determine the approaches that have and have not worked. Then, focusing on promising works that had aspects that were pertinent to the research goal and sought to evaluate (a) social inputs as sources of data and (b) court transcripts of cases that dealt with many instances of fraud.

The literature on behavioral theories was applied within specific cases involving fraudulent activity, providing insight into other characteristics that are related to threats. Empirical studies are more trustworthy when the researcher focuses on specific domains of insider threats (Schryen et al., 2016); the behavioral context was crucial to the

collected material within the present study. The research process evolved within the 11 defined steps of DSR, and introduced additional literature to support the research.

This dissertation study accurately identified the research problem in Step 2 of the research process. The research problem was based on what is known from the collected literature. The corresponding need to review additional literature can be revisited during Step 6 in the DSR process if the problem identification changes.

Also reviewed, were other input sources such as case summaries, deadlines, hearings, docket reports, filers, history documents, parties, related transactions, case review of court transcripts, plea agreements, case notes, and available court reporting in Step 3 of the DSR. A sufficient number of cases supported each of the created artifacts. The evaluation of relevancy was ensured and in alignment with the fraudulent events within Step 4 of the DSR. Data that were unrelated to the various forms of fraud were dismissed and not included in the study. The design of the IT artifact in Step 5 was a crucial research component; however, the design did require additional literature research, as identified in Step 6.

Step 7 proved to be an imperative step with DSR and withing the context of this research, allowed modifications of the research questions. For instance, one of the initial research questions left an open-ended direction for the study and was later dropped. The original DSR utilized a hypothesis refinement for this step; however, this research opted to replace the hypothesis refinement with refinement of the research questions.

During Step 8, various tools provided the processing of data, such as Garner's (1995) WEKA, Jünger and Keyling's (2013) FacePager, custom-written scripts, and the other tools listed in the resource requirement's section of this study. WEKA was used to

analyze content from Twitter, with keyword analysis from extracted data within the court transcripts relating to fraud.

The goal was to conduct as many experiments as needed to conclude with logical predictions when ranking the negative lexicons found within fraudulent court data, and then use sentiment analysis as an approach to score specific threats (Biswas, Mukhopadhyay, & Gupta, 2018). For instance, the totality of negative sentiments (see Figure 3) is evaluated by an organization's limit with scoring insider treat predictions; each organization sets their own acceptable scores.

This dissertation study examined behavioral theories and paired specific theories as applicable through observation and interpretation of court documents during Step 9. Many behavioral theories came into focus; however, six were notable and one theory appeared most prevalent across all lab experiments.

Implementation of created artifacts requires organizations to access prospective new hires' publicly available social media content. There did not appear to be ethical issues in accessing openly available data. Although, Lupton and Michael's (2017) study presented ethical issues, as participants were often highly aware that companies such as Facebook and Google track their preferences, habits, and the content they upload to social media.

The output from all tools was delivered in the form of a report that explained the findings within the insider threat predictions (see Figure 3). These findings indicated the level of threat based on data specifically centered around fraud. The scores represent the totality of all negative sentiment scores. An example of the predictions was drawn from each respective lab experiment results (-4.77, -10.45, -11.69, -16.06, -13.09, and -12.20).

All lab results demonstrated low to excessive content associated with fraud. A logical conclusion would include not hiring individuals with scores that exceed an organization's specific threshold.

Insider Threat Predictions

Score	Expected Results
0	*No Insider Threat Conditions Detected*
-5	Low Content Associated with Fraud Dectected
-10	
-15	High Content Associated with Fraud Detected
-20	
-25	Excessive Content Associated with Fraud Detected

Figure 3. Insider threat predictions.

In Step 10, this dissertation study provided a method for further research that may influence the final step, and presented a summary of the findings in Step 11. All advancements within the study significantly supports design science knowledge base and becomes a part of the knowledge contribution. Organizational leaders who have a richer understanding of insider threats can help mitigate insider threats using a preemptive approach. Preemptive approaches can improve preemployment vetting.

Instrument Development and Validation

This dissertation study highlighted synopses with instrument development and validation with granularity; listed subsequent sections. The study outcome provided output to support sentiment analysis research associated with fraud and the ranking of social media data. The instrument development and validation supports the development

of a tool to improve employee vetting and will help organizations work towards another mitigation strategy against the insider threat.

Collection of Court Documents

The aim of this research was to develop an instrument that consists of creating IT artifacts from multiple sources of data. The core instruments come from the analysis of case data of fraud through a software suite of tools listed in the resource requirement's section of this study.

This dissertation study examined prosecuted fraud cases to extract negative sentiment with the belief that uniqueness of undesirable traits from criminal activities are advantageous to the study. Understanding the courts' search parameters from the legal component led to starting the process to gain access to preceding types of documents, following two established processes for collecting court documents: (a) created an account and completed the application for multi-court exemption from the Judicial Conference's Electronic Public Access (EPA) and (b) submitted the form through regular mail delivery or electronically. No charges incur while conducting research once the fee waiver is granted; however, nominal transaction charges are incurred if fees are not waived. According to PACER (2020), nominal transaction charges are based on the following:

> The PACER cost is $0.10 per page with a cap of $3 per document, except transcripts. What is the cost for using CM/ECF? There is no additional fee associated with the CM/ECF system. Public Access to Court Electronic Records (PACER) is an electronic public access service that costs **$0.10** per page. (PACER.gov, 2020, p. 1)

A valid login and password are required after successful account creation. In addition, the client code of "SME" is required, and then used the required credentials to log in to USCourts.gov's PACER system. There was a requirement to register for accessing data through the ECF/SMG portal. Although the overall process is effortless,

the time required to complete the process necessitated a brief delay and lasted several days.

A selection to the electronic case files and PACER case locator is required, and allows logging in. It is critical to choose the appropriate court system; the chosen court system should be consistent with the EPA request. This dissertation study used the California Northern District Court and The Superior Court of California County of Santa Clara to gain public access to electronic court records. Preliminary data for initial queries consisted of the following: (a) case numbers, (b) case status (all), (c) file date and last entry date (empty), (d) nature of suit, and (e) cause of action (18:1030 computer fraud and abuse). The executed query ran after entering the above selections. PACER returned a sample listing of cases, and each case could be selected using the case number (see Table 2).

Table 2
PACER Sample Inquiry

Case number	Filed under	Dates files	Nature of suit
1001	Acme v. Individual A	Filed 01/31/2000 closed 6/13/2000	18:1030(18:1030 Computer Fraud and Abuse Act)
1002	Smarts v. Individual B	Filed 06/16/2000 closed 12/07/2000	18:1030(18:1030 Computer Fraud and Abuse Act)
1003	Taxi Co. v. Individual C	Filed 09/06/2000 closed 12/07/2000	18:1030(18:1030 Computer Fraud and Abuse Act)
1004	Jobs Inc. v. Individual D	Filed 03/19/2001 closed 10/25/2002	18:1030(18:1030 Computer Fraud and Abuse Act)
1005	Toys Co. v. Individual E	Filed 05/16/2002 closed 05/05/2004	18:1030(18:1030 Computer Fraud and Abuse Act)
1006	Cars Co. v. Individual F	Filed 09/13/2002 closed 12/19/2002	18:1030(18:1030 Computer Fraud and Abuse Act)
1007	Homes Inc. v. Individual G	Filed 03/04/2003 closed 09/17/2003	18:1030(18:1030 Computer Fraud and Abuse Act)

All counsel and parties accessing documents filed with the court are responsible for redacting personal identifiers from all downloaded documents (USCourts.Gov, 2019). The redaction of personal identification is pursuant to a Local Civil Rule 7(c)(2) and Local Criminal Rule 47(c)(2). All extracted data will have data redaction in-place and exhibit a level of anonymity to demonstrate compliance.

Each query is saved into a PDF or comma-separated values (CSV) document using a naming convention. The file's prefix uses the initial data source as its designation. For instance, PACER uses "PACER" for all electronic court records, and state and local governments used either Santa Clara, Illinois, Albemarle, Fairfax, Loudoun, or Prince William county to represent the general district courts. The second designation is the case

number followed by a numerical designation to reflect an IT artifact's association. The Bear File converter is used to convert from TEXT or CSV if files are not in the required format. The goal is to provide a repeatable document conversion process. In many instances, PDF files are opened with iSkySoft and saved into text format. Later, the files are loaded into Excel, where the content is sorted and duplicate words are removed. In the present study, multiple test cases were saved and the following file name formatting was used: "State-CaseID.TXT" (e.g., CA-C123456.txt). In the example given, the filename represents case data from California, and the identification assigned by the court is C123456. In addition, each filename was unique to each lab experiment.

Read Court Documents into WEKA

The first validation process required making the primary data accessible within WEKA and validated court case data using WEKA's explorer option to import the CSV (or .XLS) file. Then, saving the file as an attribute-relation file format (ARFF) American National Standards Institute compatible file was required. This dissertation study then created sentiment analysis by using a classification process based on sentiment weights. The initial classification used various algorithms within the lab experiments and a filtering mechanism to support supervised learning.

The initial results yielded a successful validation (see Figure 4); however, the process became a component in the pairing with other data used in the research. The actual weights were obtained by importing VADER's negative sentiment lexicons. VADER's lexicons perform exceptionally well in the social media domain (Hutto & Gilbert, 2014) and were advantageous to the present study. However, the way VADER (see Appendix B) data imports into WEKA, a custom script is used to rewrite each

sentiment's weight; thus, these scripts use the actual weight of each lexicon rather than WEKA's default values of 1.0. The only exception to this logic is to assign WEKA's default weighted values for words that do not exist in the base lexicon word list. The value of 0 is assigned in these instances.

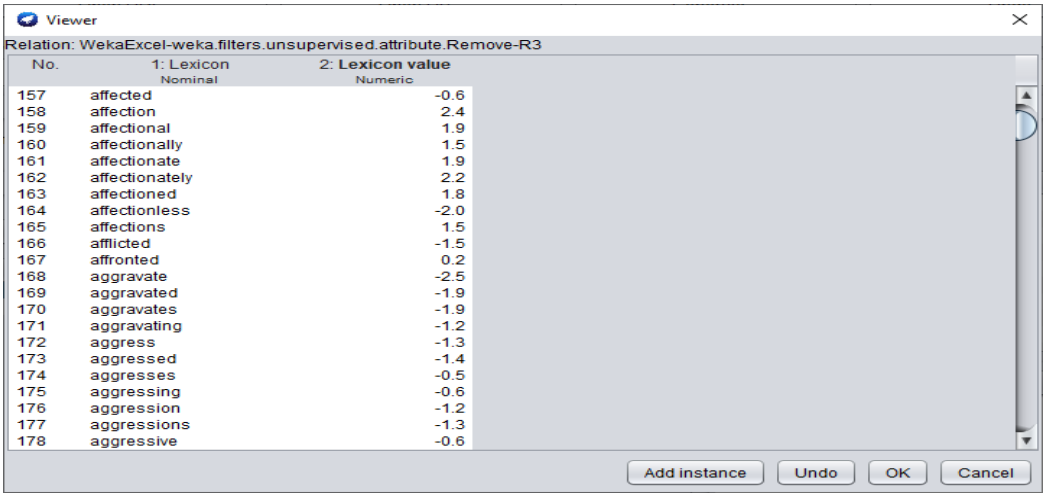


Figure 4. WEKA preprocessing readable sample data.

Produce Social Media Exports in FacePager

The second validation process required each social media export to pass through preprocessing in Microsoft Excel to (a) eliminate duplicate words using an internal sort function and (b) eliminate content deemed not relevant in building a lexicon before saving. As part of the preprocessing to ensure all data is readable and capable of being interpreted in WEKA, all validations must be met. Similarly, with court cases, an import to WEKA is performed through the program's explorer option. WEKA is expected to read the data and produce content (see Figure 4) that could be used in further analysis.

The other input source came from collecting publicly accessible Twitter tweets within FacePager. To accessed this input source required opening the FacePager

application and select the “New Database” option. In the “Save As” field, entering the database name will represent the source of Internet data (i.e., Twitter-1000), left the “Tags” field empty, and changed “Where” to a folder location (i.e., DISS901-3\Twitter). On the lower portion of the screen, “Twitter” was selected. Under the same general section, required selection of a “Resource” and required that the field was set to “/statuses/user_timeline;” this can also be set by clicking on the API image at the top portion of the screen. Within the API interface, a requirement of the same Twitter API is was applied. Other required settings included (a) adding a setup query string consisting of “https://api.twitter.com/1.1” into the base path, (b) setting the resource to “/statuses/user_timeline,” setting the parameters to “user_id and references <Object ID>,” and setting the maximum pages to 1000. Under settings, ensured “Select All Nodes” was selected, that “Maximum Errors” was set to 99, and that “Log All Requests” was checked. Using the custom table columns window, adding a new key called “text” and applied to the “Apply Column Setup” saved the setting. The “text” showed within the upper objects and query window.

Next, each value (see Figure 5) was deleted in the access token and access token secret and required secure consumer keys (see Figure 6) from the Twitter developer’s account and access was granted after Twitter vetted the application. The access settings were applied to the consumer keys (see Figure 7) within the “Authentication Settings” window. To set the access token and access token secret, required logging into Twitter and selecting “Add Node” on the upper part of the screen. This dissertation study used Calculator.net’s (2020) random number generator to create a comprehensive list of numbers from a lower limit of 1067092933653692416 and an upper limit of

1267092933653692415 and generated 20,000 unique numbers, and then sorted the numbers in ascending order. The long integers corresponded to possible Twitter IDs. Using Notepad, output was saved to a folder used in later reference (i.e., C:\Research\Twitter).

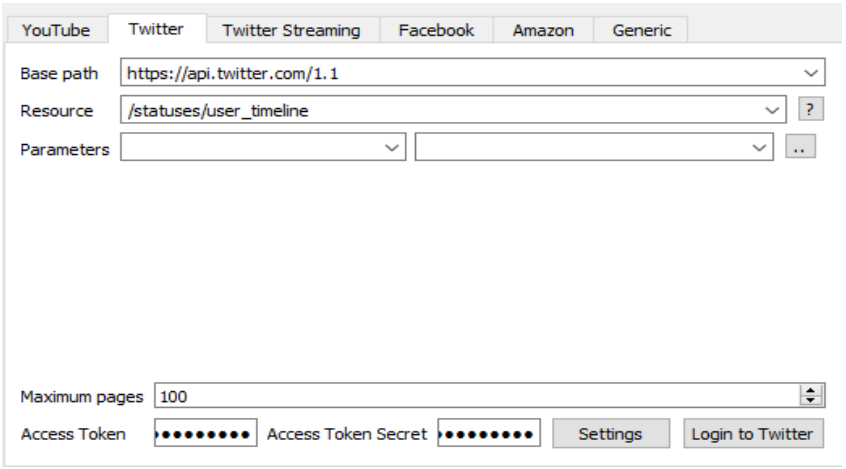


Figure 5. FacePager access token.

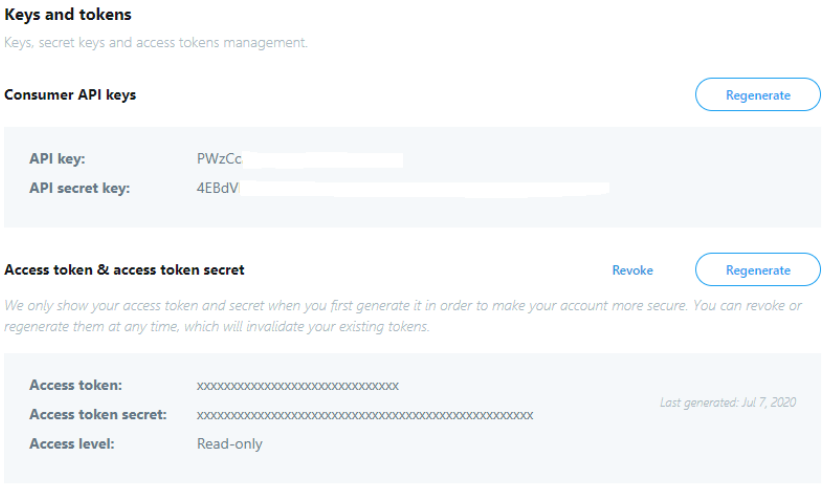


Figure 6. Consumer keys.

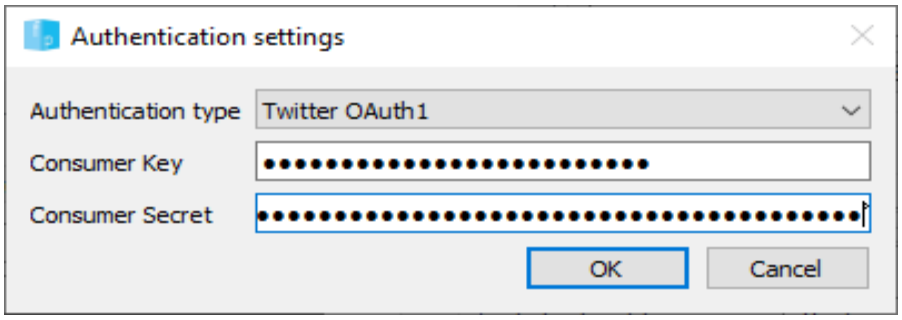


Figure 7. Authentication settings.

Next, content was copied and pasted from the Notepad into the "Add Node" box. Under "Settings," select both "All Nodes" and the "Fetch Data" options are selected. Next, highlight any object type labeled "Data" within the upper leftmost box and verify that the rightmost box had Tweets in the Key (text) field's value. Once multiple data fields showed meaningful data, all Tweets are exported by selecting all results and using use the "Export Data" option. Content was saved within the root folder under C:\ Research Twitter Tweets\TwitterData.csv.

Later, the dissertation study used Vicinitas' (2020) services to query Twitter feeds for March 2019; this query included 20,000 random tweets. Data were provided in an Excel worksheet, and only the text of the Tweets was saved to a similar file, such as C:\Research \Tweets \Vicinitas\Tweets\TwitterData.csv. Additionally, the data was saved as an Excel file from the .xlsx format to the .xls for compatibility with WEKA's Excel converter 1.0.7. This dissertation study used data from Vicinitas because of the quickest data availability. Both FacePager or Vicinitas processes works, but the latter is considered the most productive; providing faster overall processing.

Import Social Media into WEKA

The preceding section detailed how to move data into WEKA as part of the validation process; the same methodology must be applied to the FacePager data. The validation of FacePager data remains consistently the same, aside from differences in data attributes. A decision was made to save imported social media data to an ARFF compatible file (see Appendix C). Similar to court case data, there was a need to create sentiment analysis using the same word classifications. This dissertation study initially used Naïve Bayes for word classification. Naïve Bayes is a probabilistic method that has high accuracy and performance in text classification (Sari, Kurniawati, Prayitno, & Irfangi, 2019) with a probabilistic classifier for the normal distribution to model numeric attributes (Amin & Habib, 2015). However, manually traversing the classifications within WEKA led to an extensive delay while testing various classifications. Ultimately, using Auto-WEKA provided an easier initial approach, especially when dealing with large datasets, repetitive classification testing and knowing how to choose among the dozens of machine learning procedures implemented in WEKA and each procedure's hyperparameter settings to achieve good performance (Kotthoff, Thornton, Hoos, Hutter, & Leyton-Brown, 2017). Because of Auto-WEKA's ability to select the best classification, efforts later ran manual classifications comparisons against Auto-WEKA, and used Trees Random Forest, Naïve Bayes, and DMNB. The results yielded successful validations, which became a component in the pairing with other data used in the research.

Perform Sentiment Analysis

Sentiment analysis was needed to best understand the existence of fraud within the context of social media and performed in WEKA through importing and pairing each court case involving fraud with one social media input as ARFF compatible file, and then applied WEKA's Auto-WEKA document classification (see Appendix D) to select the best algorithm; the best algorithm is typically Trees Random Forest, Naïve Bayes, or DMNB. This dissertation study leveraged further filtering and analysis by using the WEKA package AffectiveTweets. The importance of using AffectiveTweets is centered around the package's ability to analyze tweets. According to Branz and Brockmann (2018), AffectiveTweets provides many capabilities and notable filtering to demonstrate negative sentiment with emphasis to anger, disgust, fear; all of which is based on sentiment lexicon scoring. In additions, AffectiveTweets can analyze social media sentiment and is key in the distant supervision method for models using unlabeled tweets (Bravo-Marquez, Frank, Pfahringer, & Mohammad, 2019), which was a critical element within this study.

Technological Aspects

The technological aspects of this study were purely driven by the tools used within the study and the ease of data collections in leveraging Internet data. Specifically, the use of the Internet played an instrumental role in the research. The Internet created the method to reach into the court systems, and more importantly, social inputs. Both endpoints provided a deeper understanding of common sentiments extracted from fraudulent cases and helped to potentially identify insider threats.

Behavioral Theory-Based Aspects

The social component of behavioral theories played an essential role in understanding the sentiment analysis' outcome. An abundant number of behavioral theories represent individuals' mindsets and explains individuals' pathway to becoming a threat. Social and internal mechanisms play a critical role in a person's mind, as these mechanisms silence one's urge to follow moral obligation when people act illegally (Shi, Booth, & Simon, 2017). Various behavioral theories were examined and became an integral component of the study.

Sample Data

This dissertation study obtained sample data from two key sources: (a) court documentation from The Superior Court of California County of Santa Clara, The Northern District of California San Francisco Division, The United States District Court Northern District of Illinois, and The United States District Court Southern District of California and (b) publicly available Twitter tweets.

The research followed the data analysis as referenced in the GPFLE section for consistency and ensured all processed data could be read through various programs. Initially, all court data was scrubbed for blank records, duplicates, and other evidence of insufficient data. In some other cases, data was not able to be read until installing OCR plug-ins. A series of custom scripts and programs corrected formatting issues and saved into a format compatible with WEKA. The collection of Twitter tweets included information requiring redaction. In these instances, user id, user screen name, bio, location, following, followers, and favorites were all excluded while exporting data into WEKA.

Through additional scripts came the resulting ARFF compatible files used by WEKA in the analysis and tweet classifications. Outside of WEKA's purview left the behavioral theory interpretations, which augmented the study's classification and analysis. The subsequent data analysis, the GPFLE process, and lab experiments provide a closer look into the lab experiments' data interactions.

Data Analysis

Sentiment analysis may be one of the best tools for making predictions within social media content. Within the context of this study, sentimental analysis was applicable and built on lexicon data from fraudulent cases. Building dictionaries for each word’s contextual characteristics—such as its order in the text, part of speech, cooccurrence with other words, and other contextual characteristics specific to the text in which the word appears—is essential (Shapiro, Sudhof, & Wilson, 2018).

A host of tools to predict insider threats were readily available. For instance, this dissertation study used each respective set of files within WEKA to perform sentiment analysis using each dictionary and each crawl of social media with FacePager. The dissertation study’s goal was to rank sentiments associated with fraud, yet conform to a lexicon-based negative social media sentiment, as Hutto and Gilbert (2014) demonstrated in their multiple-domain model. The lexicon construction’s foundation included using cases of fraud, social media inputs, and news articles, which presented a wide latitude of sentiments and were used throughout all lab experiments.

Sufficient literature on insider mitigation strategies exists; however, other avenues must be traversed, which provided a starting point for this study. An attempt was made to (a) use the baseline of negative sentiment lexicons, (b) examine criminal cases of fraud,

(c) determine similarities, (d) traverse Twitter tweets in an attempt to find Tweets that shared commonalities, (e) perform sentiment analysis, and (f) correlate common themes of behavioral theories to the outcome.

The dissertation study's data analysis provided insight into the correlation of lexicon data, increased understandings of social inputs through machine learning, promoted IT artifact creations and usage, and used Hutto and Gilbert's (2014) lexicons to analyze custom scripts and WEKA's AffectiveTweets package to examine all sources of court data to produce a summary of results.

Formats for Presenting Results

The dissertation study intended to examine social inputs coming from fraudulent-related activities and use sentiment analysis as one of its core tools. The results provided a better understanding of insider threat mitigation, as examined within employee vetting. Supplemental results revealed annotations from insider threat rankings and social-behavioral theories to support the findings.

This study's results came through outputs of classifications, IT artifact correlations of sentiment from social inputs, and insider threat lexicons from fraud. The output from DSR provided contributions to solution and problem-domain maturities by supportive means to adaptations and provide a pathway of improvements within the scope of insider threat mitigations. The presented results include charts, figures, tables, and other visual means to describe the findings, with supplemental information in the appendices and references.

Resource Requirements

Due to the nature of the research with leveraging content from social media, institutional review board approval was not required because the investigation did not directly interface with people. The information collected was publicly available from social media websites, and the content is retrieved through no special means aside from using tools designed to collect data. This dissertation study was granted a Twitter developer's account and passed Twitter's vetting process. In addition to social media data, requests for court-filed insider threat case information directly related to fraud within Albemarle, Fairfax, Loudoun, and Prince William County general district courts was initiated. This dissertation study's preliminary requests for information indicated that it is possible to obtain this information using only a case number and, in some instances, a nominal fee for court transcripts. Online information is available to the public without any special access; however, court documents fall under the Freedom of Information Act.

This dissertation study used the following hardware and software within this study. Aside from Internet connections, all hardware and software are considered operating within one physical location. Both Microsoft Windows 10 (1909) and Apple macOS High Sierra 10.13.6 operating systems provided the base for which all research tools operated and detailed in the following section.

Research Tools

WEKA

WEKA is used as an input and output tool that reads inputs from social media imported Excel data, provides output with both sentiment analysis, and uses classification output and sentiment filtering. In addition, WEKA uses a common file format to store its

data sets, thus presenting the user with a consistent view of the data regardless of what machine learning scheme may be used (Garner, 1995).

FacePager

FacePager is used for fetching publicly available data from YouTube, Twitter, and other websites based on APIs and web scraping. All data are stored in an SQLite database and may be exported to CSV (GitHub.com, 2020). FacePager was used to test fetching Twitter data during the pre-research phase. This process provided insight into methods used later within the study.

Auto-WEKA

Auto-WEKA makes it easy for non-experts to find the best classification algorithm within WEKA along with a good hyper-parameter configuration for a given application scenario, with little human time and within a reasonable amount of fully automated computation (Thornton, Hutter, Hoos, & Leyton-Brown, 2012). The classification algorithm selection process necessitated very little input. However, this study leveraged the WEKA generic object editor to set optional parameters within the Auto-WEKA classification selection process.

Auto-WEKA's classification parameters are as follows: batch size (100) debug (false), check capabilities (false), memory limit (8192), metric to optimize (error rate), best configs (1), decimal places (2), parallel runs (1), seed (123), and time limit (60). The only settings changed for the lab experiments are memory and time limit. Each respective default values of the two settings were 1024 and 15 (see Appendix D, Figure D5).

The Auto-WEKA package uses an input tool to analyze data, determines the optimal classification algorithm, and apply the algorithm to data. The process to achieve

the optimal algorithm is driven by the data being ingested. According to Kotthoff, Thornton, Hoos, Hutter, and Leyton-Brown (2017), not every classifier will apply to every dataset. The primary reason is due to the classification process's ineffectiveness in handling missing data and applies a subset of classifiers within the optimal solution. Examples of the different classifiers are outlined within the lab experiments, with Auto-WEKA selecting either Random Forest, Naïve Bayes, or DMNB. However, Random Tree, Lazy IBk, and SVM were also evaluated.

AffectiveTweets

AffectiveTweets is implemented as a package for WEKA machine-learning workbench and provides methods for calculating state-of-art affect analysis features from tweets (Bravo-Marquez et al., 2019). This dissertation study used AffectiveTweets to analyze social media content and produce output in determining sentiment scores; detailed within the lab experiments.

WEKAExcel

WEKAExcel is used to import Excel files into WEKA. Once successful imports occur, WEKA can save data in a standard file format that becomes usable to any machine-learning scheme.

Java

Java 8u261 is a requirement for the WEKA version used in the research.

Visual Studio

Visual Studio 2019 v16.8.4 was used to create custom code (see Appendix E, F, G, H, and I).

Office 365

Office 365 was used to produce research reports.

Google Scholar

Google Scholar was used to collect research content separate from Nova Southeastern University's library.

PACER

The PACER system has brought the citizens ever closer to the courthouse. Public access to court documents is faster, better, and cheaper than at any prior time in U.S. history (Martin, 2008). PACER is a useful input tool in requesting court data.

iSkySoft Editor

iSkySoft editor was used to read input PDF court data and to create exportable files used by custom scripts within the study.

NSU Library

The Nova Southeastern Library was instrumental in the collection of literature. In many instances, literature did not exist without paying for subscriptions to many of the top journals.

FiOS Internet

FiOS Internet provided communications to allow the research to take place.

Summary

The methodology overview detailed the 11 steps of DSR, including creating the IT artifacts and ranking the sentiment analysis of multisource data by analyzing each collected data set. Each data set represented a single IT artifact. In addition to the social and fraud inputs, behavioral theories contributed towards a holistic representation of each

IT artifact. This chapter outlined a listing of required resources that were needed to support data collections, analysis, and suggested sampling. The instrument development and validation process provided a glimpse into pairing data, performing sentiment analysis, ranking potential insider threats, and why specific traits for fraud takes place.

Chapter 4

Results

Data Analysis

All lab experiments correlated to the creation of each IT artifact and provided a way to rank fraudulent case data in the sentiment of Twitter content. The data analysis revealed the attributes associated with insider threats and identified various machine-learning algorithms that can be used to leverage inputs from court documents on fraud. The data provided ample content to construct the subsequent IT artifacts and pairing of each artifact with the negative VADER sentiments. Similarly, the study's uniqueness included specific VADER's negative sentiments that were extracted from court proceedings (see Appendix J) and deemed relatable to criminal activities.

Initially, this dissertation study planned to use Naïve Bayes as the primary classifier; however, the discovery of Auto-WEKA and manual WEKA, appeared to be more beneficial than any specific algorithm. For instance, Auto-WEKA supports an intelligent algorithm selection that is beyond other manual selecting of classifications within WEKA. Auto-WEKA supports the selection of 30 classifications by automatically reviewing performance generalizations and applies model optimizations previously believed to be only a manual process (Kotthoff, Thornton, & Hutter, 2017).

The output from each experiment is carefully examined in the subsequent sections, with emphasis placed upon each unique court case. Additionally, the processes included discoveries with negative sentiments found within social inputs, specifically within Tweets. Each IT artifact is traceable to a specific case of fraud, and relations to negative sentiments are uncovered within tweets. The analysis coincides with many of the referenced behavioral theories to illustrate the social aspect of the IT artifact.

General Procedures for Lab Experiments (GPFLE)

All lab experiments required three data sources: (a) court documentation retrieved through USCourts.gov's website, (b) negative social media sentiment, and (c) Twitter tweets. The uniqueness of each artifact was established through individually collected court data. All other data collections were replicated through each experiment to include over 7,000 of Hutto and Gilbert's (2014) negative sentiment lexicons and 20,000 Twitter tweets from 2019. For ease of repetition, the experiments referenced GPFLE as a starting point.

The sequential effort began with collecting data from the courts. This process required the study to convert the source PDF files into text by using the iSkySoft document converter. Next, saving the output file (e.g., CA-C1903821.txt) into a text file with each line within the file representing each word in the referenced case number. A few instances did require using the optical character recognition (OCR) plug-in. In addition, this dissertation study created a custom program called PDF-Text-ToProcessedText (see Appendix I) to read each line of the converted text file and write a new file that included corrections for formatting. Other processing included importing the text file to Microsoft Excel, sorted and removed duplicate words by ensuring the

data/sorting advanced unique records was the only option selected, saved the Excel file back into the text file, and saved the text file into the .xls format to support compatibility with WEKA's Excel converter 1.0.7. Additional preparations required launching WEKA from a command prompt within the c:\program files\WEKA\subdirectory, and then invoking the "Java -Xmx8192m -jar WEKA.jar" command. Failure to change the application's access to additional heap memory led to heap errors, errors when marshaling XML response, and unanticipated program termination. A later discovery narrowed down the preceding problem with Auto-WEKA and using the graphical user interface on the Apple OSX. Because of the issue, launching WEKA from the command line as annotated above, was the only course of action in allowing the automatic classification process to continue.

The collection of Twitter data was received as an Excel file and then saved to a CSV file. The last of the initial parallel effort concluded with converting Hutto and Gilbert's (2014) social media VADER sentiment lexicons into a readable format for later processing. The dissertation study accomplished the preceding by using a custom program to pair negative sentiments within the court data.

The dissertation study developed a sequence of short programs to stage and test the data. The assignment of weights to court lexicons (see Appendix E) program reads the text files, creates a new output file with the content from the input, and paired with associate sentiment weight from Hutto and Gilbert's (2014) sentiment weights. The output is read into WEKA, along with Twitter data, and sentiment analysis is performed. Converting Twitter tweets into fixed words (see Appendix F) was accomplished by reading the 20,000 tweets from the master CSV file and convert into plain text. Two

sequencers (see Appendix G and H) first read each line from court documents and looks up occurrences within the matching file. If a match is found, the output is written as a hit; with matching lexicon and weight (i.e., trouble, -1.5). If a match is not found, the output is written to the same file with a neutral zero weight. The second sequencer reads each line from the hits file and locates possible hits in the Twitter file. If matching tweets exist, record the lexicon and number of times found (i.e., trouble, -1.5). If a match is not found, the output is written to the same file using a zero weight as 0.0. Because iSkysoft created an unformatted layout, the last program converted iSkysoft's output into formatted text (see Appendix I).

Using the previously created "hits" file as an input, examining the Tweets for words that existed in the "hits" file and recorded the word and total occurrences within the Tweets was needed. Next, the final output file was saved as "Found-In-Social-Media." The file was later used by Excel and prepared by the following steps:

- Opened the "Found-In-Social-Media" file using Excel.
- Used the Excel Import Wizard to select "Delimited with Tabs"
- Reviewed the data in the lower portion of the window, then clicked "Next" followed by "Finished"
- Saved with "Save-as CSV" (MS-DOS; *.csv)

After the programs produced the initial data captures, this dissertation study used WEKA to load the "Found-In-Social-Media.csv" file by opening and then saving the CSV fie to an ARFF file. Works by Bravo-Marquez et al. (2019) led to the use of AffectiveTweets text classification 1.0.2 filter for analyzing the sentiments of Tweets. The package supporting the WEKA/filters/supervised/attribute and made was installed to

support Tweet selections, and used the preprocess tab under attributes to create the statistics for 28 attributes.

Lastly, the ending component in the artifact creation required examining relevant behavioral theories that applied to the court-collected data. The dissertation study's understanding of the context of the events leading up to the fraud warranted an examination of the behavioral aspects that contributed to the study's conclusion.

Lab Experiment 1: Artifact 1

The artifact ID identified by the first lab experiment was SC-1903821. In this instance, Santa Clara, California provided data from a fraud case that was estimated at $500,000 in losses to a local business. The following steps were executed to create the artifact, apply both court and reporting notes, and connect to a specific behavioral theory. This dissertation study followed the GPFLE process for Lab Experiment 1 and supplemented the process with the following activities: determine artifact scores, tweet negative emotion scores, optimal classification, and artifact negative summary.

WEKA became an instrumental tool of the study, and within the process tab, provided various scores (see Table 3) for 27 attributes. The attributes included values derived from "TweetToInputLexiconFeatureVector," "TweetToLexiconFeatureVector," "TweetToWordListCountFeatureVector," and "TweetToSentiStrengthFeatureVector" filters. The Artifact 1 detailed scores can be found in Table 3. The most prominent Tweet negative emotion (see Appendix K) scores (see Figure 8) are the four listed within the negative summary and the same presented in an alternate view (see Figure 9).

Table 3
Artifact 1 Detailed Scores

Attributes	Minimum	Maximum	Mean	StdDev
Sentiment weight	-2.5	2.4	.986	.911
Hits	0	90	27.886	26.452
NRC-Affect-Intensity-Anger-Score	0	.485	.008	.063
NRC-Hash-Sent-posScore	0	.573	.069	.187
NRC-Hash-Sent-negScore	-1.244	0	-.268	.183
NRC-10-Anger	0	1	.017	.13
NRC-10-Trust	0	1	.143	.351
NRC-10-Negative	0	1	.023	.15
NRC-10-Positive	0	1	.253	.441
NRC-10-Expanded-Anger	0	.037	.004	.011
NRC-10-Expanded-Anticipation	0	.079	.009	.023
NRC-10-Expanded-Disgust	0	.01	.001	.003
NRC-10-Expanded-Fear	0	.01	.001	.003
NRC-10-Expanded-Joy	0	.094	.009	.028
NRC-10-Sadness	0	.016	.002	.005
NRC-10-Surprise	0	.037	.004	.011
NRC-10-Trust	0	.02	.002	.006

Table 3
Artifact 1 Detailed Scores (continued)

Attributes	Minimum	Maximum	Mean	StdDev
NRC-10-Expanded-Negative	0	.106	.01	.031
NRC-10-Expanded Positive	0	.121	.012	.036
SentiWordnet-posScore	0	.354	.082	.124
SentiWordnet-negScore	-.344	0	-.115	.109
Mpqa-posCount	0	1	.36	.481
Mpqa-negCount	0	1	.17	.13
BingLiu-negCount	0	1	.17	.13
AFINN-posScore	0	2	.623	.875
AFINN-negScore	-3	0	-.051	.344
S140-posScore	0	.27	.172	.06
S140-negscore	-.18	0	-.003	.023

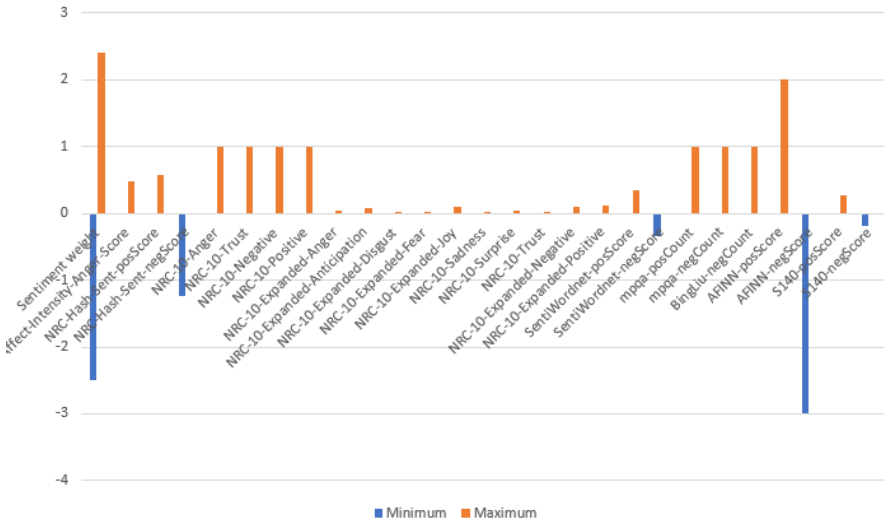


Figure 8. Artifact 1 Tweet negative emotion scores.

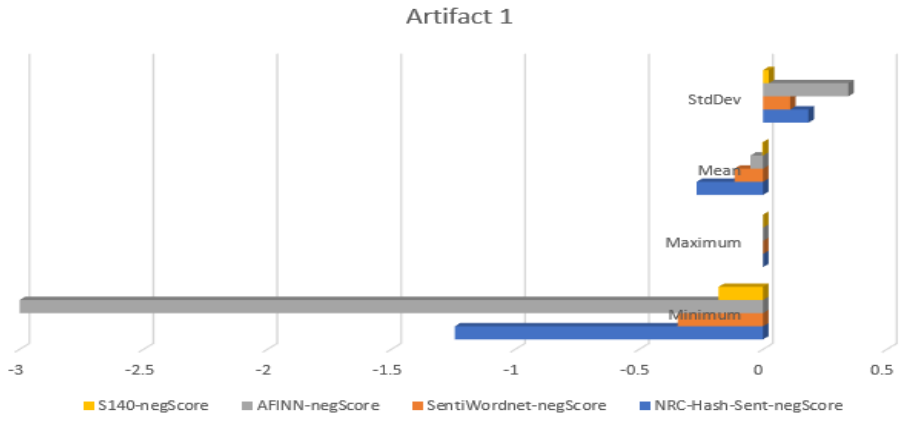


Figure 9. Artifact 1 negative summary.

The Naïve Bayes classification was initially thought to be the preferred classifier within all lab experiments; however, after manually testing various classifiers, the Trees Random Forest classification algorithm was selected and then applied against court data (see Figure 10). Although the algorithm provided comparable accuracy over some of the other classifiers within this lab experiment, not all experiments produced similar results; in this instance, Trees Random Forest (see Appendix L) yielded a conducive Kappa score (see Appendix M) of 0.8055.

```
Correctly Classified Instances         137               87.2611 %
Incorrectly Classified Instances        20               12.7389 %
Kappa statistic                          0.8055
Mean absolute error                      0.0531
Root mean squared error                  0.1642
Relative absolute error                 24.8382 %
Root relative squared error             51.5841 %
Total Number of Instances              157

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Lexicon
                 0.643     0.077     0.450       0.643     0.529     0.484     0.944     0.462     active
                 1.000     0.000     1.000       1.000     1.000     1.000     1.000     1.000     bold
                 0.000     0.000     ?           0.000     ?         ?         0.500     0.019     complaint
                 0.000     0.000     ?           0.000     ?         ?         0.500     0.006     felony
                 1.000     0.000     1.000       1.000     1.000     1.000     1.000     1.000     justice
                 1.000     0.054     0.954       1.000     0.976     0.950     0.973     0.954     number
                 0.214     0.035     0.375       0.214     0.273     0.232     0.936     0.426     parties
Weighted Avg.    0.873     0.039     ?           0.873     ?         ?         0.962     0.851
```

Figure 10. Artifact 1 Trees Random Forest classification.

The following section evaluates the outcome from the classification process and delves deeper into the True Positive Rate (TPR), False Positive Rate (FPR), and the Precision. First, according to Azar, Elshazly, Hassanien, and Elkorany (2014), the True Positive Rate (TPR) is representative of positive classes correctly classified by a model is achieved. Second, the False Positive Rate (FPR) represents the fraction of negative classes that are identified as positive. And third, the Precision is the exactness of a classifier (Kaur & Saini, 2015), and ideally having a higher precision closer to 1.0 correlates to fewer false positives.

Within the context of this experiment, the TPR's weighted average came to *0.873* and considered acceptable when comparing a perfect rate of 1.0. In contrast, the FPR's weighted average scored a *0.039*; acceptable and low. The non-weighted average for the Precision was manually calculated at an average of 0.539 and appears comparable to other experiments, though lower than expected.

Although this experiment produced TPR and FPR numbers without values, this was likely due to a sparse data capture. Instances of zero-valued classes accounted for two negative lexicons; *complaint* and *felony*. Class values greater than zero and less than one accounted for two lexicons; *active* and *parties* are not listed with negativity. And the last class with values equal to one accounted for three lexicons *bold*, *justice*, and *number*; not appearing in negative lexicon listings (see Appendix Q). Furthermore, this experiment's classification of zero-valued TPR classes could be identifiers with lexicons associated with fraud.

The recall is defined as the number of relevant items retrieved as a proportion of all the relevant items that might potentially be retrieved (Walters, 2016) and in the

instance from this experiment, two classifications with two lexicons yielded no recall. While the F-Measure effectively references the TP to the arithmetic mean of predicted positives and real positives, being a constructed rate normalized in an idealized value (Powers, 2020), several instances was undefined and is believed to be due zero TP values within the limited data capture.

Other classification details include the Matthews Correlation Coefficient (MCC). The MCC is a more reliable statistical rate which produces a high score only if the prediction obtained good results (Chicco & Jurman, 2020). As with the Precision, Recall, and F-Measure, the best explanation for MCC's unknown values is likely due to the comparatively small size of both positive and negative records within the data. When '?' symbols appear in the output, the specific class may have not enough samples or none of the samples can be assigned to the class (Stackoverflow, 2021). With the increase in data within subsequent experiments, it is postulated the numbers will not remain undefined. Lastly, the examination of the Receiver Operating Characteristic (ROC), a plot of the true positive rate against the false-positive rate at various threshold settings (Egieyeh, Syce, Malan, & Christoffels, 2018) had a weighted average of 0.962. According to Fan, Upadhye, and Worster (2006), general interpretation of the value is high discriminatory, yet anything higher and closer to 1.0 is very rare. Interestingly, the following experiments' volume of data significantly increases, and class accuracies are further examined with the expectancy to help postulate cohesions shared within all experiments.

Lastly, the ending component in the artifact creation required the examination of relevant behavioral theories as applied to the court collected data. According to D'Addona (2019):

> Over the course of his 30 years in the coaching profession, the defendant had the opportunity to train, develop, and establish swimmers and programs throughout the United States. He came to SCSC in September of 1995 as the Associate Head Coach, under Dick Jochums. Working with Jochums, SCSC went on to win two National titles for men and one national title combined. Following the retirement of Jochums in December of 2006, the defendant was elevated to the position of Head Coach of Santa Clara Swim Club. In that position, the club saw tremendous growth in terms of both performance and number of swimmers and programs offered to the swimming community. In September of 2009, the defendant was also given the title of CEO of the organization. (p. 1)

The turn of unfortunate events shared by D'Addona aligns with the RAT. In this instance, the offender leveraged his tenure and promotions through the ranks to place himself in a position with access to financial components within the organization. According to Cohen & Felson's (1979), definition of RAT, the circumstances surrounding the embezzlement was demonstrated through the lack of capable guardians against criminal activities, and the offender sought suitable targets; all in alignment with the theory.

In summary, the data from the fraudulent case SC-1903821 demonstrated the lowest below zero AFINN-negScore (see Appendix K) value of -3 when evaluating 90 hits from comparing court case lexicons from the case and Twitter tweets having like sentiments. In the same analysis, the AffectiveTweets collected the negative SentiWordnet, resulting in the automatic annotation of all the synsets of WordNet with notations of "positivity," "negativity," and "neutrality" (Baccianella, Esuli, & Sebastiani, 2010). In these instances, the tweet's focus was placed upon negativity of -0.344, signifying a minor threat. In comparison, the AFINN score yielded -3 and represented words with a score that ran between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment (Silge & Robinson, 2020). Moreover, the NRC attributes were derived from word-level emotion association lexicon

for about 14,200 word types (Mohammad & Turney, 2013), which produced the lowest recording with a HASH-SENT-negScore of -1.244.

Additionally, the identified behavioral theory added value when joined as part of a comprehensive approach through the experiments. The outcome produced a common theme shareable throughout the research, which allowed identifying areas of weakness in a collective effort to detect the insider threat.

Lab Experiment 2: Artifact 2

The second lab experiment was identified by artifact ID 320-CR-00266. Unlike the first experiment, a significant increase of court data was included. The United States District Court, Northern District of California, San Francisco Division provided data from a case with 23 counts to scheme, artifice to defraud investors, and estimated in the millions of dollars. According to Anderson (2020), the case involved engagement in illegal activities relating to false financial statements, abetting, bank fraud, and wire transfer fraud.

The following steps were executed to create the artifact, apply both court and reporting notes, and connect to a specific behavioral theory. The distinguishing and contrasting attributes beyond the first experiment came through additional data. Lab Experiment 2 offered a 5-fold data increase. This dissertation study followed the GPFLE process for Lab Experiment 2 and supplemented with the following activities: determine artifact scores, tweet negative emotion scores, optimal classification, and artifact negative summary. The detailed scores for Artifact 2 can be found in Table 4.

Table 4
Artifact 2 Detailed Scores

Attributes	Minimum	Maximum	Mean	StdDev
Sentiment weight	-2.8	1.8	.247	1.006
Hits	0	91	23.122	25.461
NRC-Affect-	0	.394	.033	.104
NRC-Hash-Sent-	0	1.066	.098	.227
NRC-Hash-Sent-	-4.99	0	-.229	.51
NRC-10-Anger	0	1	.096	.295
NRC-10-Trust	0	1	.201	.402
NRC-10-Negative	0	1	.096	.295
NRC-10-Positive	0	1	.21	.408
NRC-10-Expanded-	0	.033	.004	.008
NRC-10-Expanded-	0	.198	.014	.034
NRC-10-Expanded-	0	.007	.001	.002
NRC-10-Expanded-	0	.021	.004	.008
NRC-10-Expanded-	0	.112	.013	.027
NRC-10-Sadness	0	1	.044	.205
NRC-10-Surprise	0	0	0	0
NRC-10-Trust	0	1	.201	.402
NRC-10-Expanded-	0	1	.096	.295
NRC-10-Expanded	0	1	.21	.408
SentiWordnet-	0	.175	.01	.027
SentiWordnet-	-.291	0	-.083	.08
Mpqa-posCount	0	1	.004	.066
Mpqa-negCount	0	1	.105	.307
BingLiu-negCount	0	1	.105	.307
AFINN-posScore	0	2	.672	.489
AFINN-negScore	-1.174	0	-.328	.923
S140-posScore	0	.927	.279	.291
S140-negScore	-.18	0	-.085	.274

As in the previous experiment, WEKA continued to be an instrumental tool for the study. The process tab provided various statistics for 27 attributes. The study used the preprocess tab and created statistics for 27 attributes (see Table 4). The attributes included values derived from “TweetToInputLexiconFeatureVector,” “TweetToLexiconFeatureVector,” “TweetToWordListCountFeatureVector,” and “TweetToSentiStrengthFeatureVector” filters. The most prominent tweet negative

emotion (see Figure 11) scores are the four listed within the negative summary (see Appendix K), and the same presented in an alternate view (see Figure 12).

Like Experiment 1, the study initially thought Naïve Bayes was the preferred classifier within all lab experiments. However, after testing various classifiers, the study selected the Trees Random Forest classification algorithm and applied this algorithm against court data (see Figure 13), which provided a slightly lower score when compared to the previous experiment, yet not a substantially lower percentage.

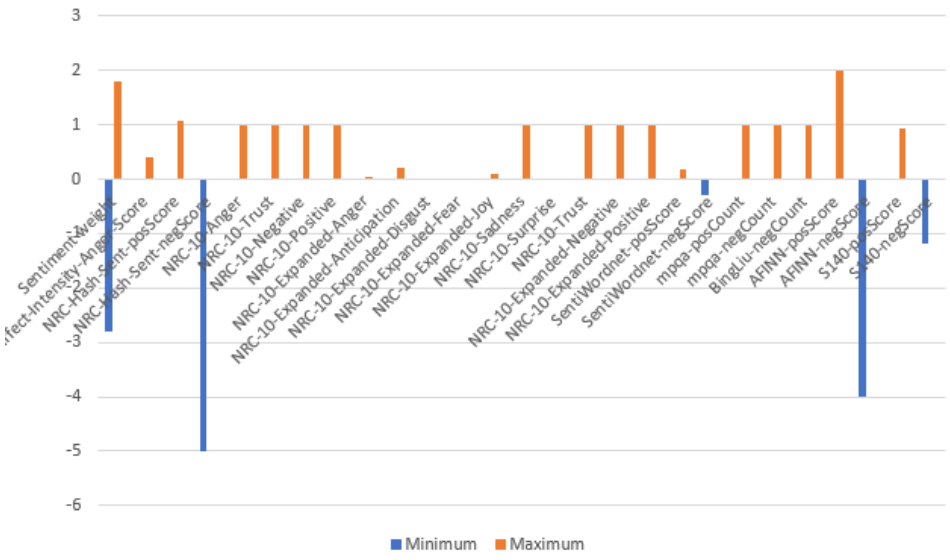


Figure 11. Artifact 2 Tweet negative emotion scores.

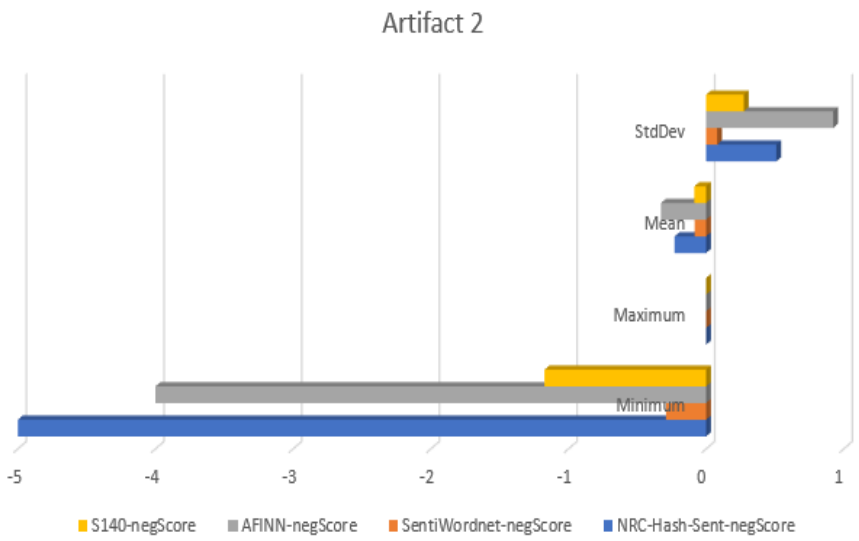


Figure 12. Artifact 2 negative summary.

```
=== Summary ===

Correctly Classified Instances         175               84.9515 %
Incorrectly Classified Instances        31               15.0485 %
Kappa statistic                          0.8106
Mean absolute error                      0.0362
Root mean squared error                  0.1407
Relative absolute error                 30.316  %
Root relative squared error             58.2385 %
Total Number of Instances              206

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Lexicon
                 0.000    0.000    ?          0.000   ?          ?        0.500     0.010     assets
                 1.000    0.005    0.955      1.000   0.977      0.974    1.000     0.998     created
                 0.778    0.036    0.500      0.778   0.609      0.603    0.975     0.479     creating
                 0.000    0.000    ?          0.000   ?          ?        0.500     0.015     credits
                 0.000    0.000    ?          0.000   ?          ?        0.500     0.058     fraud
                 1.000    0.011    0.889      1.000   0.941      0.938    0.999     0.979     hide
                 1.000    0.011    0.917      1.000   0.957      0.952    0.997     0.961     legal
                 0.000    0.000    ?          0.000   ?          ?        0.500     0.010     liability
                 1.000    0.061    0.400      1.000   0.571      0.613    0.970     0.400     limited
                 1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     matter

                 0.000    0.000    ?          0.000   ?          ?        0.500     0.010     promises
                 0.125    0.010    0.333      0.125   0.182      0.185    0.961     0.377     shares
                 0.000    0.000    ?          0.000   ?          ?        0.500     0.005     substantial
                 1.000    0.027    0.783      1.000   0.878      0.873    0.975     0.725     united
Weighted Avg.    0.850    0.009    ?          0.850   ?          ?        0.940     0.797
```

Figure 13. Artifact 2 Trees Random Forest classification.

Within the context of this experiment, the TPR's weighted average came to *0.850* and considered acceptable when comparing a perfect rate of 1.0. In contrast, the FPR's weighted average scored a *0.009*; exceptionally low. The non-weighted average for the Precision was manually calculated at an average of 0.999 and appears considerably higher than experiment one.

Although this experiment produced TPR and FPR numbers without values, this is best explained by the class not having enough samples. Instances of zero-valued classes accounted for two negative lexicons; *liability* and *fraud*. Class values greater than zero and less than one accounted for no negative lexicons. And the last class with values equal to one accounted for two lexicons *limited*, and *hide*; appearing in negative lexicon listings (see Appendix R). Furthermore, this experiment's classification of both zero-valued and one-valued TPR classes might be identifiers with lexicons associated with fraud.

The recall is defined as the number of relevant items retrieved as a proportion of all the relevant items that might potentially be retrieved (Walters, 2016) and in the instance from this experiment, two classifications with two lexicons yielded no recall. While the F-Measure effectively references the TP to the arithmetic mean of predicted positives and real positives, being a constructed rate normalized in an idealized value (Powers, 2020), with several instances undefined and was believed to be due zero TP values within the limited data capture.

As in the preceding experiment, other classification details include MCC and F-Measure. Comparing to the previous experiment, the best explanation for MCC's unknown values was likely due to the comparatively small size of both positive and negative records within the data. With the increase in data within subsequent experiments

three through six, again it is postulated the numbers will not always remain as undefined. Lastly, the examination of the ROC had a weighted average of 0.940. Fan, Upadhye, and Worster's (2006) general interpretation of the value is high discriminatory, and again; anything higher and closer to 1.0 is very rare. Interestingly, the following experiments' volume of data significantly increases, and class accuracies are further examined with the expectancy to help postulate cohesions shared within all experiments.

Lastly, the ending component in the artifact creation required the examination of relevant behavioral theories as applied to the court collected data. The following excerpt provides an in-depth description on the case and Justice.gov (2020) stated the following:

> The complaint and information, the defendant, 36, of San Francisco, California, is alleged to have orchestrated multiple schemes to defraud his victims. The defendant founded a venture capital company that he used between 2013 and 2016 to raise and manage four annual funds whose purpose was to invest in start-up companies, and particularly companies in the field of virtual reality technologies.
>
> The information filed today alleges that the defendant partially funded his capital commitment to the second of those funds by committing bank fraud. Specifically, in 2014, the defendant made false statements about his wealth to his bank while refinancing his home mortgage and while obtaining a $300,000 personal loan, and poured some of the ill-gotten money he obtained from the bank into the second of his funds.
>
> In 2015, the information alleges that the defendant took excess money in venture capital fees from one of the funds he was raising and managing, and therefore faced a shortfall at the end of the year that he did not wish to report to his investors. At the end of 2015, the information alleges that the defendant engaged in a scheme to defraud a bank by making false statements and misrepresentations to the bank in order to obtain a $4 million line of credit to pay back the fund from which he had taken excess fees. In so doing, the defendant attempted to deceive his investors into believing the fund was well-managed and was following the operating agreements the investors understood controlled the management of the fund.
>
> In February 2016, according to the allegations laid out in the information, the defendant engaged in a scheme to defraud an investor with respect to a $2 million investment that it believed it was making directly into a virtual reality content production company that the defendant contended he wholly-owned. It is alleged that, rather than using that investment as he had represented, the defendant used most of it for other purposes.

> The complaint then alleges that, in July 2016, the defendant engaged in a scheme to defraud as many as five separate investors when he induced them to wire a total of $1.35 million under the premise of investing in the untraded stock of a privately-held software company. The complaint charges the defendant with knowingly engaging in a scheme to defraud one investor by representing to that organization that its money would be used to purchase the software company's shares. According to the complaint, on the same day the money was wired, the defendant took the money from the bank account designed to make the investment and sent it to a main operating bank account, from which it was used for many purposes. The complaint alleges that no stock in the software company was ever purchased.
>
> Finally, the information sets out allegations about a series of investors as to whom the defendant engaged in a scheme to defraud in 2015 and 2016 by inducing their investments in his managed funds under the premise he would use the money for investments in "frontier edge" technologies and take only certain limited fees for the management of the funds. Instead, the defendant took more fees than to which he was entitled and invested far less of the money he raised than the operating agreements disclosed to the investors contemplated.
>
> Today's allegations in the criminal complaint and information state that the evidence has established that since 2013, the defendant fraudulently obtained at least $18.8 million through his illegal conduct. (p. 1)

After reviewing the court's news release and reviewing the case as filed with the courts from June 26, 2020, this particular case appeared to align with two behavioral theories: the RAT and the TPB. As Cohen and Felson (1979) posited, unlawful activities are brought together through conditions exhibited in this case, along with investors (the targets) and lacked protectors to these types of criminal activities. Equally, the TPB demonstrates the insufficiency of following any type of behavioral control, antecedents of attitudes, subjective norms, and perceived behavioral control that lead to predictors with intentions and actions (Ajzen, 1991).

In these two instances with behavioral theories, the offender targeted investors who were all naive to activities outside of their purview. Furthermore, the offender used an elaborate strategy of being in control of many financial schemes that lacked any type of checks and balances. The lack of behavioral norms accepted by society did not play a role, which explains why TPB is in alignment with the outcome.

Lab Experiment 3: Artifact 3

The third lab experiment was identified by Artifact ID 320-CR-00245. Similar to the last experiment's volume of court data, the lexicons retrieved from this case yielded an 18% hit rate increase over the previous experiment. The United States District Court, Northern District of California, San Francisco Division provided data from a case with multiple counts, including defrauding 21 financial institutions under false pretenses and bank fraud.

Unlike previous experiments, a data conversion from OCR to PDF required an additional step, which was outlined within the GPFLE process. This dissertation study followed the same GPFLE process during Lab Experiment 3 and supplemented the process with the following activities: determine artifact scores, tweet negative emotion scores, optimal classification, and artifact negative summary.

This dissertation study used the WEKA preprocess tab and created statistics for 27 attributes (see Table 5) to include attributes with the lowest (see Appendix K) Tweet negative emotion sentiment scores (see Figure 14). The attributes included values derived from "TweetToInputLexiconFeatureVector," "TweetToLexiconFeatureVector," "TweetToWordListCountFeatureVector," and "TweetToSentiStrengthFeatureVector" filters. An alternate view is presented (see Figure 15). Unlike prior experiments and through manual testing various classifiers, Discriminative Multinomial Naïve Bayes (DMNB) was the selected classification. A concerted effort included manually selecting other classifiers, but none fared better than DMNB when applied against the Twitter tweets (see Figures 16–19). More importantly, the classifier correctly identified 94.79%, and held a Kappa score of 0.9449.

Table 5
Artifact 3 Detailed Scores

Attributes	Minimum	Maximum	Mean	StdDev
Sentiment weight	-3.2	3.1	.712	1.542
Hits	0	329	71.032	82.094
NRC-Affect-Intensity-Anger-Score	0	.882	.033	.133
NRC-Hash-Sent-posScore	0	5	.25	.482
NRC-Hash-Sent-negScore	-4.999	0	-.382	.454
NRC-10-Anger	0	1	.064	.245
NRC-10-Trust	0	1	.106	.307
NRC-10-Negative	0	1	.122	.327
NRC-10-Positive	0	1	.209	.407
NRC-10-Expanded-Anger	0	.713	.026	.069
NRC-10-Expanded-Anticipation	0	.314	.063	.089
NRC-10-Expanded-Disgust	0	.464	.012	.035
NRC-10-Expanded-Fear	0	.74	.024	.08

Table 5
Artifact 3 Detailed Scores (continued)

Attributes	Minimum	Maximum	Mean	StdDev
NRC-10-Expanded-Joy	0	.724	.108	.217
NRC-10-Sadness	0	.807	.32	.119
NRC-10-Surprise	0	.149	.025	.043
NRC-10-Trust	0	.684	.118	.18
NRC-10-Expanded-Negative	0	.956	.076	.169
NRC-10-Expanded Positive	0	.883	.207	.304
SentiWordnet-posScore	0	1.539	.335	.476
SentiWordnet-negScore	-1.067	0	-.094	.221
Mpqa-posCount	0	1	.512	.5
Mpqa-negCount	0	1	.147	.354
BingLiu-negCount	0	1	.151	.358
AFINN-posScore	0	3	.795	.99
AFINN-negScore	-4	0	-.477	.938
S140-posScore	0	1.707	.32	.381
S140-negscore	-2.148	0	-.194	.407

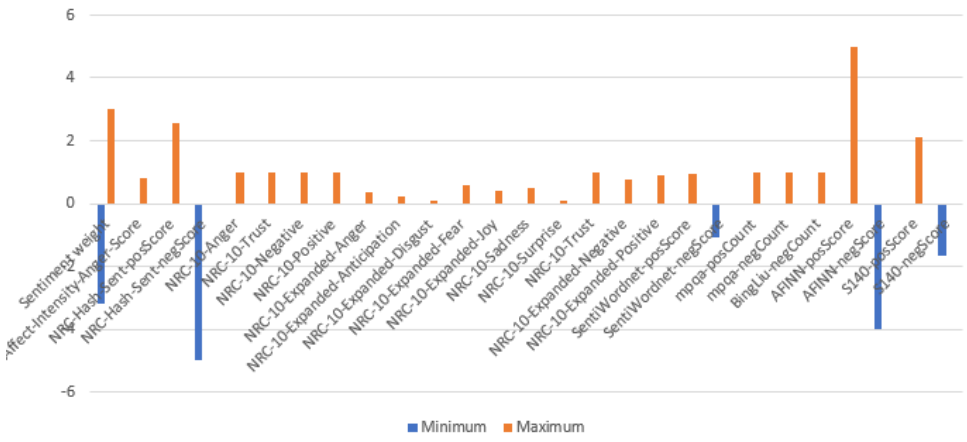


Figure 14. Artifact 3 Tweet negative emotion scores.

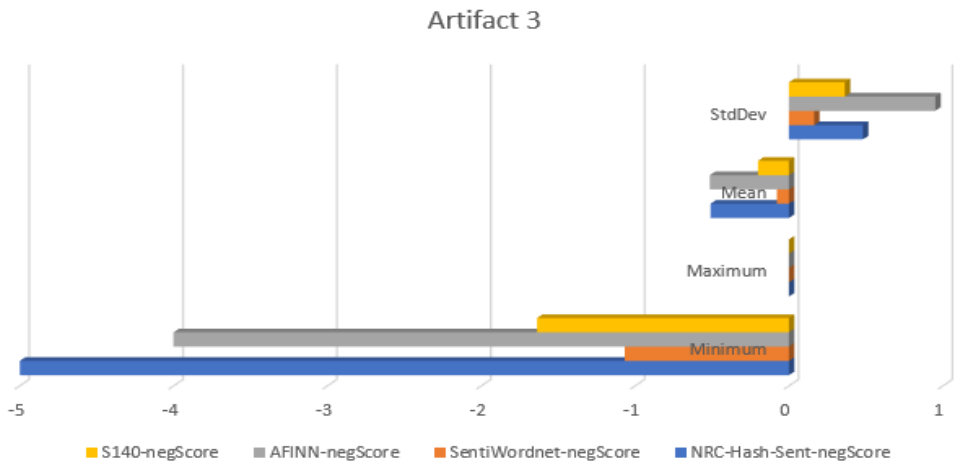


Figure 15. Artifact 3 negative summary.

Correctly Classified Instances	1750	94.7996 %
Incorrectly Classified Instances	96	5.2004 %
Kappa statistic	0.9449	
Mean absolute error	0.0082	
Root mean squared error	0.0494	
Relative absolute error	34.1997 %	
Root relative squared error	45.0781 %	
Total Number of Instances	1846	

Figure 16. Artifact 3 Discriminative Multinomial Naïve Bayes classification - Part 1.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	ability
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	abuse
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	accept
0.000	0.000	?	0.000	?	?	0.883	0.009	acceptance
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	accepted
1.000	0.000	1.000	1.000	1.000	1.000	1.000	0.988	accepting
1.000	0.002	0.833	1.000	0.909	0.912	1.000	1.000	agreement
1.000	0.001	0.946	1.000	0.972	0.972	1.000	1.000	allow
0.000	0.000	?	0.000	?	?	0.978	0.078	approval
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	approved
0.167	0.000	1.000	0.167	0.286	0.408	1.000	0.948	arrested
0.000	0.000	?	0.000	?	?	0.994	0.367	assets
1.000	0.001	0.889	1.000	0.941	0.943	1.000	1.000	benefit
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	commit
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	committed
0.400	0.000	1.000	0.400	0.571	0.632	1.000	1.000	conspiracy
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	credit
0.000	0.000	?	0.000	?	?	0.999	0.792	credits
1.000	0.001	0.857	1.000	0.923	0.925	1.000	1.000	crime
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	criminal
1.000	0.001	0.952	1.000	0.976	0.976	1.000	1.000	dangerous
1.000	0.003	0.829	1.000	0.906	0.909	1.000	1.000	dear
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	death
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	defense
0.000	0.000	?	0.000	?	?	0.950	0.011	determination
0.000	0.000	?	0.000	?	?	0.995	0.242	determined
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	effective
0.250	0.000	1.000	0.250	0.400	0.500	1.000	1.000	engaged
0.100	0.000	1.000	0.100	0.182	0.315	0.997	0.598	entitled
0.667	0.000	1.000	0.667	0.800	0.816	1.000	1.000	error
0.000	0.000	?	0.000	?	?	0.961	0.014	felony
1.000	0.002	0.949	1.000	0.974	0.973	1.000	1.000	fine
1.000	0.001	0.963	1.000	0.981	0.981	1.000	1.000	fit

Figure 17. Artifact 3 Discriminative Multinomial Naïve Bayes - Part 2.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	fraud
0.000	0.000	?	0.000	?	?	0.807	0.003	gains
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	grant
0.000	0.000	?	0.000	?	?	0.694	0.002	granting
1.000	0.001	0.889	1.000	0.941	0.943	1.000	1.000	gross
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	guilty
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	honorable
0.000	0.000	?	0.000	?	?	0.978	0.044	imposed
0.000	0.000	?	0.000	?	?	0.962	0.026	improvements
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	injury
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	interest
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	justice
1.000	0.002	0.969	1.000	0.984	0.983	1.000	1.000	leave
0.000	0.000	?	0.000	?	?	0.985	0.119	liability
1.000	0.001	0.969	1.000	0.984	0.984	1.000	1.000	loss
0.000	0.000	?	0.000	?	?	0.987	0.061	losses
1.000	0.004	0.870	1.000	0.931	0.931	1.000	1.000	low
0.000	0.000	?	0.000	?	?	0.998	0.658	mandatory
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	matter
1.000	0.003	0.948	1.000	0.973	0.972	1.000	1.000	number
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	offense
0.000	0.000	?	0.000	?	?	0.828	0.003	offenses
1.000	0.003	0.833	1.000	0.909	0.911	1.000	1.000	original
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	outstanding
1.000	0.004	0.708	1.000	0.829	0.840	1.000	0.984	parties
1.000	0.001	0.991	1.000	0.996	0.995	1.000	1.000	pay
1.000	0.012	0.943	1.000	0.971	0.966	1.000	1.000	please
0.000	0.000	?	0.000	?	?	0.988	0.081	promises
0.000	0.000	?	0.000	?	?	0.984	0.062	questioned
0.000	0.000	?	0.000	?	?	0.998	0.702	recommended
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	risk
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	safety
1.000	0.002	0.778	1.000	0.875	0.881	1.000	1.000	secure

Figure 18. Artifact 3 Discriminative Multinomial Naïve Bayes - Part 3.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
	0.000	0.000	?	0.000	?	?	0.960	0.030	secured
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	sentence
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	smart
	1.000	0.001	0.968	1.000	0.984	0.983	1.000	1.000	special
	1.000	0.006	0.935	1.000	0.966	0.964	1.000	1.000	support
	0.000	0.000	?	0.000	?	?	0.966	0.175	suspended
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	united
	0.964	0.006	0.730	0.964	0.831	0.836	0.996	0.649	victim
	0.000	0.000	?	0.000	?	?	0.993	0.362	victims
	0.000	0.000	?	0.000	?	?	0.946	0.010	violation
	0.000	0.001	0.000	0.000	0.000	-0.002	0.991	0.286	violence
	0.000	0.000	?	0.000	?	?	1.000	1.000	weapon
	0.000	0.000	?	0.000	?	?	0.992	0.144	wells
Weighted Avg.	0.948	0.003	?	0.948	?	?	0.999	0.964	

Figure 19. Artifact 3 Discriminative Multinomial Naïve Bayes - Part 4.

Within the context of this experiment, the TPR's weighted average came to *0.948* and considered acceptable when comparing a perfect rate of 1.0. In contrast, the FPR's weighted average scored a *0.003*; exceptionally low. The non-weighted average for the Precision was manually calculated at an average of 0.958 and appears comparable to experiment two.

Although this experiment produced TPR and FPR numbers without values, this is likely due the inability to be assigned to the class, and not based the volume of data; opposite of the previous experiment. Instances of zero-valued classes accounted for eleven negative lexicons; *felony*, *imposed*, *liability*, *losses*, *offenses*, *questioned*, *suspended*, *victims*, *violation*, *violence*, and *weapon*. Class values greater than zero and less than one accounted for four negative lexicons; *arrested*, *conspiracy*, *error*, and *victim*. And the last class with values equal to one accounted for fifteen lexicons *abuse*, *crime*, *criminal*, *dangerous*, *death*, *fraud*, *gross*, *guilty*, *injury*, *leave*, *loss*, *low*, *offense*, *pay*, and *risk*; appearing in negative lexicon listings (see Appendix S). Furthermore, this experiment's classification of all TPR classes might be identifiers with lexicons associated with fraud and differ from the preceding experiments.

The takeaway from using DMNB within the experiment is the classifier provided similar results for TPR classes having zero values, yet provided similar benefits when

seeking negative social lexicons that could be associated with fraud. Moreover, DMNB demonstrated an extremely low FPR and a high percentage of precision when properly evaluated; not applicable across the board when both TPR and FPR values were zero. Conjointly, the zero values for TPR and FPR impacted both F-Measure, and MCC, while the weighted average for ROC was 0.999. This dissertation study did not expect these results, yet it is posited the outcome is based on some data not being assigned to a class; notably from previous experiments, the same held true. However, some lexicon class attributes are conducive in vetting social media tweets and could help mitigate insider threats from the most earliest onset.

Lastly, the ending component in the artifact creation required the examination of relevant behavioral theories as applied to the court collected data. The following article provides a summary of events. Fox Business (2014) stated the following:

> A California man has pleaded guilty for his role in a nationwide automobile loan fraud scheme the U.S. Secret Service discovered in Pennsylvania last year. The defendant, 30, of Hercules, California, allegedly solicited straw purchasers and then lied about their creditworthiness so the conspirators could obtain auto loans that were never paid back. The crooks split the money among themselves, costing 21 victim banks and credit unions in several states $1 million to $2.4 million, Assistant U.S. Attorney Marshall Piccinini told a federal judge Friday. Among other things, the defendant faked borrowers' tax and wage documents and used vehicle identification numbers from real cars that were not actually for sale. Banks were told the loans were being used to pay for vehicles being sold by two fictitious firms, Gold Coast Group Worldwide and AM Auto Groups. Banks lost money because there was no real collateral to secure the loans. Philip pleaded guilty before U.S. District Judge David Cercone in Pittsburgh because the scheme was uncovered last year by U.S. Secret Service agents in Erie, about two hours north of Pittsburgh. The Erie Federal Credit Union and Erie Community Credit Union were among the financial institutions victimized, Piccinini said. The prosecutor wouldn't say how many other people prosecutors believe were involved in the scheme, only that the defendant is the first to be prosecuted. Piccinini wouldn't say if others would be charged in Pennsylvania or other jurisdictions where the fraud played out. In all, 64 phony borrowers attempted 150 bogus loans. Piccinini wouldn't say how many loans were successful, or how many people in the scheme worked as "brokers," "managers," or "processors" of

> loans described in the charges filed against the defendant. The defendant established bank accounts in the name of a phony firm, -- Investments Inc., through which some of the loan proceeds were moved. Investigators have traced $544,000 from the scheme to the defendant's bank accounts, but say at least $219,000 of that was paid out to other participants. Piccinini wouldn't say what may have happened to the rest of the money. The government is not seeking to force the defendant to forfeit any money, which is often done in financial fraud crimes, but he may be ordered to pay restitution to the banks when he's sentenced Feb. 23. The defendant pleaded guilty to bank fraud and a separate count of conspiracy to commit bank and wire fraud. Both charges carry up to 30 years in prison. (p. 1)

Review of the court's case filed on June 19, 2020 appeared to align with two behavioral theories: the RAT and the TPB. As in the previous experiment and applicable is Cohen and Felson's (1979) theory with unlawful activities coming together through conditions exhibited in this case; along with banks and credit unions (the targets) and lacked protectors to these types of criminal activities. Equally important, the TPB demonstrates the insufficiency of following any type of behavioral control, antecedents of attitudes, subjective norms, and perceived behavioral control that lead to predictors with intentions and actions (Ajzen, 1991), and appears to be demonstrated by greed.

In these two instances, the offenders appeared to target investors who were all blind-eyed to activities outside of their knowledge using an elaborate strategy of being in control of many financial schemes that lacked any type of checks and balances.

Lab Experiment 4: Artifact 4

The fourth lab experiment was identified by the artifact ID 4-15-CV-01490. Although similar to prior experiments, the case differed by changing from a single deceptive defendant to a business representing a different view of an entity's fraudulent activities. The experiment included an increase of volume with court data. The lexicons retrieved from the case yielded a 38% hit rate increase over the previous experiment. Like the previous experiment, data were retrieved from The United States District Court,

Northern District of California. The San Francisco Division provided data from a company that defrauded its customers with merchandise that did not meet the State's code for selling products with formaldehyde. According to USCourts.gov (2020) Northern District of California:

> The action arises from Defendants' unlawful, unfair, fraudulent and misleading advertising, marketing and selling of their Chinese-manufactured laminate flooring ("laminate flooring") to consumers in California as compliant with formaldehyde emission standards promulgated by the California Air Resources Board ("CARB"). Defendants engaged in and continue to engage in this misleading advertising campaign in an effort to deceive consumers into purchasing its laminate flooring products. In fact, the products contain and emit formaldehyde at levels in excess of CARB standards. Formaldehyde, a colorless gas, is a substance known to cause cancer. The National Toxicology Program within the U.S. Department of Health and Human Services has classified formaldehyde as a known human carcinogen. Formaldehyde exposure in the short-term can cause irritation of the skin, eyes, nose, and throat, respiratory problems such as asthma, and neurological impairment. Long-term exposure to formaldehyde can result in an increased risk of developing certain types of cancer. Children and the elderly are at a heightened risk from formaldehyde exposure. Putting profits ahead of safety, the company specifically recommends its laminate flooring for consumers who have children. "Laminate flooring is recommended for an environment where the 'challenges' of children and pets exist, and it is not significantly noisier than other hard flooring surfaces. (p. 2–3)

Just as the previous experiment required a data conversion from PDF using OCR, the same was applied to this experiment. The documents appeared to be scanned from within the courts and into an electronic format, with many pages not correctly aligned during the scanning process and required the use of OCR software.

This dissertation study continued to follow the same GPFLE process for Lab Experiment 4. This process was supplemented with the successive activities in capturing all sentiment attributes (see Table 6), with the lowest negative sentiment scores (see Appendix K) and validation using the Naïve Bayes classification. The attributes included values derived from "TweetToInputLexiconFeatureVector," "TweetToLexiconFeatureVector," "TweetToWordListCountFeatureVector," and

"TweetToSentiStrengthFeatureVector" filters. As expected, four negative sentiments scored low and yielded a score of -2.08, whereas the S140-negScore and the NRC-Hash-Sent-negScores both yielded a score of -4.999 (see Figure 20). An alternate view is also presented with the negative summary (see Figure 21).

Table 6
Artifact 4 Detailed Scores

Attributes	Minimum	Maximum	Mean	StdDev
Sentiment weight	-3.4	3.2	.424	1.006
Hits	0	1007	178.555	25.461
NRC-Affect-Intensity-Anger-Score	0	.667	.018	.104
NRC-Hash-Sent-posScore	0	1.416	.142	.227
NRC-Hash-Sent-negScore	-4.999	0	-.325	.51
NRC-10-Anger	0	1	.039	.295
NRC-10-Trust	0	1	.099	.402
NRC-10-Negative	0	1	.092	.295
NRC-10-Positive	0	1	.153	.408

Table 6
Artifact 4 Detailed Scores (continued)

Attributes	Minimum	Maximum	Mean	StdDev
NRC-10-Expanded-Anger	0	.746	.076	.008
NRC-10-Expanded-Anticipation	0	.285	.061	.034
NRC-10-Expanded-Disgust	0	.384	.067	.002
NRC-10-Expanded-Fear	0	.875	.048	.008
NRC-10-Expanded-Joy	0	.39	.053	.027
NRC-10-Sadness	0	.72	.059	.205
NRC-10-Surprise	0	.132	.027	0
NRC-10-Trust	0	.684	.087	.402
NRC-10-Expanded-Negative	0	.982	.267	.295
NRC-10-Expanded Positive	0	.883	.148	.408
SentiWordnet-posScore	0	1.539	.255	.027
SentiWordnet-negScore	-.208	0	-.184	.08
Mpqa-posCount	0	1	.552	.066
Mpqa-negCount	0	1	.152	.307
BingLiu-negCount	0	1	.126	.307
AFINN-posScore	0	3	.729	.489
AFINN-negScore	-4	0	-.432	.923
S140-posScore	0	3.614	.22	.291
S140-negscore	-4.999	0	-.372	.274

The preceding attributes represents the direct correlations between sentiment found in court cases that is paired to known negative sentiments used in the GPFLE process and content extracted from Tweets. Thus, one can hypothesize that negative sentiment associated with forms of fraud can be extracted from social media, as shown in the collective experiments' output.

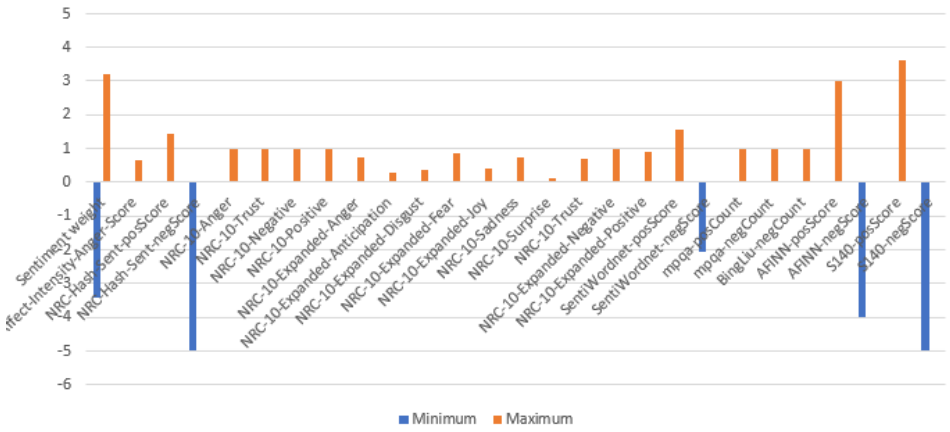


Figure 20. Artifact 4 Tweet negative emotion scores.

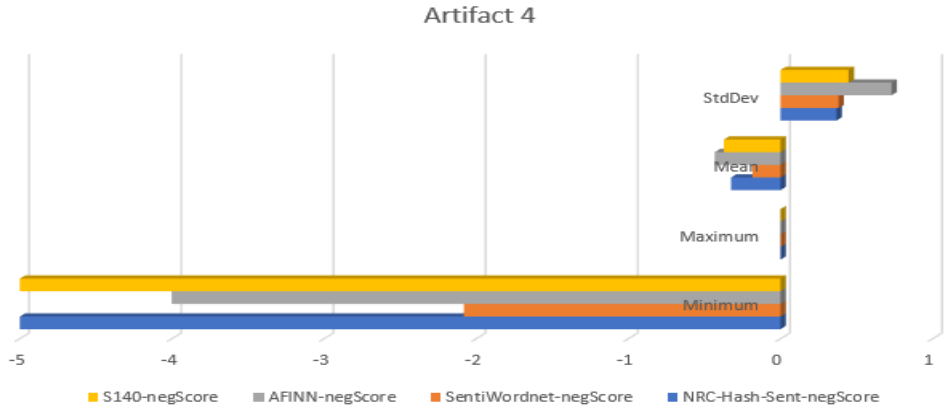


Figure 21. Artifact 4 negative summary.

The Naïve Bayes classification score was average when compared with other experiments (see Figures 22–27) and correctly identified the classification instances 70% of the time. The important takeaway was the overall average for all experiments' classification rather than this particular classification results. In this instance, instead of using the Auto-WEKA package, manually testing classifiers led to Naïve Bayes as the

best option and produced fair results with the accuracy of classifications and a Kappa score of 0.6822.

Correctly Classified Instances	4149	70.0135 %
Incorrectly Classified Instances	1777	29.9865 %
Kappa statistic	0.6822	
Mean absolute error	0.0044	
Root mean squared error	0.0486	
Relative absolute error	31.4921 %	
Root relative squared error	58.067 %	
Total Number of Instances	5926	

Figure 22. Artifact 4 Naïve Bayes classification – Part 1.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.500	0.004	0.120	0.500	0.194	0.244	0.995	0.098	acceptable
0.000	0.001	0.000	0.000	0.000	-0.000	0.846	0.001	admits
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	admitted
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.175	advanced
0.719	0.003	0.845	0.719	0.777	0.776	0.997	0.890	agree
0.000	0.001	0.000	0.000	0.000	-0.001	0.999	0.238	approval
0.000	0.001	0.000	0.000	0.000	-0.001	0.995	0.065	asset
0.000	0.002	0.000	0.000	0.000	-0.001	0.994	0.041	assured
0.800	0.000	0.923	0.800	0.857	0.859	1.000	0.926	attacks
0.692	0.001	0.643	0.692	0.667	0.666	0.999	0.793	award
0.000	0.001	0.000	0.000	0.000	-0.000	0.895	0.002	awarded
0.000	0.003	0.000	0.000	0.000	-0.002	0.993	0.094	benefits
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	best
0.000	0.001	0.000	0.000	0.000	-0.001	0.999	0.359	cancer
0.767	0.002	0.926	0.767	0.839	0.839	0.998	0.949	care
0.611	0.005	0.440	0.611	0.512	0.515	0.995	0.434	certain
0.500	0.001	0.333	0.500	0.400	0.408	0.999	0.440	challenges
0.000	0.001	0.000	0.000	0.000	-0.000	0.961	0.004	charities
0.289	0.005	0.325	0.289	0.306	0.301	0.991	0.308	clear
0.000	0.004	0.000	0.000	0.000	-0.003	0.991	0.108	commitment
0.214	0.003	0.158	0.214	0.182	0.182	0.995	0.205	committed
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.123	complained
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.102	complaint
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.125	complaints
0.333	0.002	0.091	0.333	0.143	0.173	0.997	0.113	creates
0.545	0.004	0.566	0.545	0.556	0.552	0.996	0.626	cut
0.000	0.001	0.000	0.000	0.000	-0.000	0.967	0.005	deceive
0.000	0.001	0.000	0.000	0.000	-0.000	0.981	0.009	deceived
0.000	0.002	0.000	0.000	0.000	-0.001	0.996	0.107	delay

Figure 23. Artifact 4 Naïve Bayes classification – Part 2.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
1.000	0.000	0.909	1.000	0.952	0.953	1.000	1.000	demand
0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.267	denying
0.200	0.003	0.118	0.200	0.148	0.152	0.995	0.152	desire
0.417	0.002	0.333	0.417	0.370	0.371	0.999	0.464	devastating
0.733	0.001	0.880	0.733	0.800	0.802	1.000	0.928	dream
0.000	0.002	0.000	0.000	0.000	-0.001	0.998	0.164	effectively
0.300	0.002	0.188	0.300	0.231	0.236	0.997	0.213	engage
0.000	0.001	0.000	0.000	0.000	-0.001	0.996	0.105	engaged
0.400	0.002	0.167	0.400	0.235	0.257	0.998	0.195	engaging
0.000	0.001	0.000	0.000	0.000	-0.001	0.991	0.096	ensure
0.000	0.001	0.000	0.000	0.000	-0.001	0.995	0.150	entitled
0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.267	excluded
0.600	0.001	0.429	0.600	0.500	0.507	0.999	0.388	exclusive
0.571	0.001	0.400	0.571	0.471	0.477	0.999	0.416	exposed
0.000	0.000	0.000	0.000	0.000	-0.000	0.985	0.011	extends
0.000	0.002	0.000	0.000	0.000	-0.002	0.997	0.224	fail
0.571	0.001	0.667	0.571	0.615	0.616	0.999	0.756	failed
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.132	failing
0.000	0.000	0.000	0.000	0.000	-0.000	0.979	0.008	fails
0.357	0.008	0.172	0.357	0.233	0.243	0.988	0.204	fair
0.000	0.001	0.000	0.000	0.000	-0.000	0.786	0.001	faulty
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	fight
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	fraud
0.762	0.001	0.928	0.762	0.837	0.839	0.999	0.951	giving
0.000	0.002	0.000	0.000	0.000	-0.002	0.993	0.198	greater
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	growing
0.300	0.001	0.300	0.300	0.300	0.299	0.998	0.360	guarantee
0.978	0.001	0.984	0.978	0.981	0.981	1.000	1.000	hard
0.533	0.001	0.533	0.533	0.533	0.532	0.999	0.644	harm
0.552	0.006	0.320	0.552	0.405	0.417	0.994	0.289	honest
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	illegal
0.250	0.001	0.200	0.250	0.222	0.223	0.999	0.238	immoral

Figure 24. Artifact 4 Naïve Bayes classification – Part 3.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.528	0.007	0.487	0.528	0.507	0.501	0.993	0.453	important
0.000	0.005	0.000	0.000	0.000	-0.003	0.991	0.111	increase
0.000	0.002	0.000	0.000	0.000	-0.002	0.996	0.150	increased
0.143	0.001	0.111	0.143	0.125	0.125	0.997	0.173	injured
0.667	0.001	0.444	0.667	0.533	0.544	0.999	0.385	injury
0.476	0.002	0.500	0.476	0.488	0.486	0.998	0.427	interest
0.200	0.001	0.143	0.200	0.167	0.168	0.998	0.199	interests
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.232	lawsuit
0.852	0.000	0.958	0.852	0.902	0.903	1.000	0.973	legal
0.865	0.000	0.970	0.865	0.914	0.915	1.000	0.978	lies
0.600	0.001	0.667	0.600	0.632	0.632	0.999	0.799	limited
0.724	0.014	0.487	0.724	0.582	0.585	0.990	0.493	lost
0.319	0.003	0.455	0.319	0.375	0.377	0.995	0.431	low
0.478	0.005	0.289	0.478	0.361	0.369	0.994	0.267	lower
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	lowering
0.870	0.001	0.964	0.870	0.914	0.914	1.000	0.988	matter
0.769	0.002	0.455	0.769	0.571	0.590	0.999	0.438	matters
0.000	0.000	0.000	0.000	0.000	-0.000	0.997	0.135	misleading
0.905	0.002	0.989	0.905	0.945	0.936	0.999	0.994	no
0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.227	offends
0.000	0.000	0.000	0.000	0.000	-0.000	0.967	0.005	oppressive
0.412	0.005	0.200	0.412	0.269	0.284	0.994	0.190	parties
0.465	0.014	0.325	0.465	0.383	0.378	0.984	0.308	party
0.503	0.026	0.505	0.503	0.504	0.478	0.970	0.479	please
0.577	0.001	0.652	0.577	0.612	0.612	0.998	0.731	problems
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.164	profits
0.000	0.001	0.000	0.000	0.000	-0.001	0.991	0.184	promote
0.273	0.005	0.237	0.273	0.254	0.250	0.991	0.271	protect
0.500	0.000	0.333	0.500	0.400	0.408	0.999	0.268	punish
0.000	0.001	0.000	0.000	0.000	-0.001	0.999	0.304	questioned
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.183	recommended
0.000	0.001	0.000	0.000	0.000	-0.000	0.945	0.003	recommends

Figure 25. Artifact 4 Naïve Bayes classification – Part 4.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.273	0.002	0.176	0.273	0.214	0.218	0.996	0.194	refuse
0.000	0.000	0.000	0.000	0.000	-0.001	0.996	0.111	refused
0.333	0.001	0.222	0.333	0.267	0.271	0.999	0.291	relief
0.902	0.000	0.965	0.902	0.932	0.932	1.000	0.990	respect
0.000	0.002	0.000	0.000	0.000	-0.002	0.991	0.109	responsible
0.545	0.002	0.300	0.545	0.387	0.403	0.998	0.285	risk
0.500	0.002	0.182	0.500	0.267	0.301	0.998	0.203	risks
0.659	0.001	0.794	0.659	0.720	0.721	0.999	0.865	safe
0.000	0.000	0.000	0.000	0.000	-0.000	0.895	0.002	safest
0.000	0.003	0.000	0.000	0.000	-0.003	0.996	0.235	safety
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.141	satisfied
0.000	0.002	0.000	0.000	0.000	-0.002	0.995	0.216	saved
0.125	0.001	0.200	0.125	0.154	0.157	0.998	0.253	scare
0.806	0.001	0.893	0.806	0.847	0.848	1.000	0.966	serious
0.359	0.016	0.266	0.359	0.306	0.296	0.977	0.258	share
0.333	0.002	0.182	0.333	0.235	0.245	0.998	0.218	significant
0.404	0.005	0.373	0.404	0.388	0.383	0.993	0.391	strong
0.000	0.000	0.000	0.000	0.000	-0.000	0.945	0.003	substantial
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.259	suffer
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.185	suffered
0.375	0.001	0.429	0.375	0.400	0.400	0.999	0.435	superior
0.215	0.011	0.330	0.215	0.261	0.252	0.979	0.376	support
0.258	0.018	0.426	0.258	0.321	0.305	0.968	0.458	sure
0.595	0.005	0.702	0.595	0.644	0.640	0.996	0.802	top
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	touted
0.278	0.001	0.385	0.278	0.323	0.325	0.997	0.372	trouble
0.000	0.002	0.000	0.000	0.000	-0.001	0.996	0.105	trusting
0.279	0.011	0.215	0.279	0.243	0.236	0.981	0.222	truth
0.000	0.001	0.000	0.000	0.000	-0.001	0.999	0.225	unaware
0.000	0.000	0.000	0.000	0.000	-0.000	0.996	0.058	unethical
0.000	0.000	0.000	0.000	0.000	-0.000	1.000	0.367	unfair
0.474	0.002	0.409	0.474	0.439	0.438	0.997	0.459	united

Figure 26. Artifact 4 Naïve Bayes classification – Part 5.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
	0.478	0.002	0.440	0.478	0.458	0.457	0.997	0.406	value
	0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.115	violate
	0.000	0.001	0.000	0.000	0.000	-0.000	0.971	0.006	violation
	0.000	0.000	0.000	0.000	0.000	-0.000	0.986	0.012	violations
	0.994	0.000	0.997	0.994	0.996	0.995	1.000	1.000	want
	0.000	0.000	0.000	0.000	0.000	-0.000	0.998	0.134	warn
	0.894	0.001	0.987	0.894	0.938	0.936	0.999	0.992	well
	0.986	0.000	0.986	0.986	0.986	0.985	1.000	1.000	wrong
	0.590	0.010	0.674	0.590	0.629	0.618	0.988	0.754	yes
Weighted Avg.	0.700	0.005	0.736	0.700	0.713	0.709	0.994	0.757	

Figure 27. Artifact 4 Naïve Bayes classification – Part 6.

Within the context of this experiment, the TPR's weighted average came to *0.700* and considered acceptable when comparing a perfect rate of 1.0. In contrast, the FPR's weighted average scored a *0.005*; comparable to the previous experiment. The weighted average for the Precision was 0.736 and appears lower than other experiments, but reasonable.

Again, this experiment produced TPR and FPR numbers without values; no positive data. Some instances with zero-valued classes accounted for twenty-eight negative lexicons; *cancer, complained, complaint, complaints, deceive, deceived, delay, denying, excluded, fail, failing, fails, faulty, lawsuit, misleading, offends, oppressive, questioned, refused, suffer, suffered, unaware, unethical, unfair, violate, violation, violations,* and *warn*. Class values greater than zero and less than one accounted for twenty-four negative lexicons; *attacks, cut, devastating, exposed, failed, hard, harm, immoral, injured, injury, lies, limited, lost, low, lower, no, problems, punish, refuse, risk, risks, scare, trouble,* and *wrong*. And the last class with values equal to one accounted six lexicons *demand, fight, fraud, illegal, lowering,* and *touted* (see Appendix T).

Unlike previous experiments, this experiment supplied more data; 5,926 class instances. Because of this, the total number of undefined TPR and FPR results did not exist; only zero values. In turn, the calculations for Recall, F-Measure, MCC, and ROC was possible. For these attributes, the following outcome was likely. First, the weighted averages for Recall yielded 0.700 and considered good in terms of correctly labeling lexicons. Second, the F-Measure yielded 0.713; acceptable. Third, the MCC yielded 0.709, not a complete agreement but nearest to 1.0 and acceptable. The ROC yielded 0.994 as its optimal threshold of false positives. The supportive outcome provided the identification of lexicon classes conducive in vetting social media tweets and could help mitigate insider threats from the most earliest onset.

Lastly, the impact of Artifact 4 included both business partners and consumers and appeared to fall within two theories. The theory of reasoned action (TRA) is used to reveal the meaningful effects of attitudes and subjective norms. In this particular case, the

business appeared to follow a subjective norm common with other business owners' practices. The behavior is in alignment with Hale, Householder and Greene's (2002) assessment to subjective norms. Similarly, the business specifically sought suitable targets in the absence of guardians against crime (Cohen & Felson, 1979) in order to carry out their business practices and part of the routine activity theory (RAT).

Lab Experiment 5:Artifact 5

Lab Experiment 5 was identified by Artifact I.D. 3-16-cv-02600. This experiment demonstrated an example of fraud with a deceptive business practice, posing a threat to consumers and business partners. In this instance, a corporation knowingly practiced a deceptive business model to increase its profits. A comparison of the experiment's data collection to other samples places the volume of data inline to several other experiments with 33 pages of court-produced documents.

Data were retrieved from The United States District Court, Northern District of Illinois. The Northern District provided data from a class-action lawsuit that alleged an automotive manufacturer violated the Clean Air Act and EPA guidelines through the selling of faulty vehicles as part of a much larger scheme to deceive EPA testing procedures. According to USCourts.gov (2020) Northern District of Illinois:

> Any person to manufacture or sell, or offer to sell, or install, any part or component intended for use with, or as part of, any motor vehicle engine, where a principal effect of the part or component is to bypass, defeat, or render inoperative any device or element of design installed on or in a motor vehicle or motor vehicle engine in compliance with regulations under this subchapter, and where the person knows or should know that such part or component is being offered for sale or installed for such use or put to such use. (p. 5)

This dissertation study followed the same GPFLE process in Lab Experiment 5 and supplemented the process with the successive activities in capturing all sentiment attributes (see Table 7) with the lowest tweet negative emotion scores (see Figure 28) and

validation using the OneR classification. The attributes included values derived from "TweetToInputLexiconFeatureVector," "TweetToLexiconFeatureVector," "TweetToWordListCountFeatureVector," and "TweetToSentiStrengthFeatureVector" filters. All provided the following results and expanded with further granularity. Analogous to previous experiments, four negative (see Appendix K) sentiments scored low, producing an NRC-Hash-Sent-negScore of -4.999. Equally, the SentiWordnet-negScore (-1.696), S140-negScore (-2.41), and AFINN-negScore (-4) was comparable to other experiments. An alternate view (see Figure 29) is also presented and shows the lowest four negative sentiment score summary.

Table 7
Artifact 5 Detailed Scores

Attributes	Minimum	Maximum	Mean	StdDev
Sentiment weight	-2.8	3.2	1.019	1.391
Hits	0	1771	344.527	428.429
NRC-Affect-Intensity-Anger-Score	0	.75	.007	.053
NRC-Hash-Sent-posScore	0	3.117	.231	.39
NRC-Hash-Sent-negScore	-4.999	0	-.27	.413
NRC-10-Anger	0	1	.017	.129
NRC-10-Trust	0	1	.161	.367
NRC-10-Negative	0	1	.057	.232
NRC-10-Positive	0	1	.22	.414
NRC-10-Expanded-Anger	0	.686	.096	.116
NRC-10-Expanded-Anticipation	0	.404	.051	.086
NRC-10-Expanded-Disgust	0	.728	.146	.185
NRC-10-Expanded-Fear	0	.727	.111	.153
NRC-10-Expanded-Joy	0	.724	.083	.184
NRC-10-Sadness	0	.807	.07	.108
NRC-10-Surprise	0	.149	.032	.035
NRC-10-Trust	0	.472	.065	.106
NRC-10-Expanded-Negative	0	.985	.337	.383

Table 7
Artifact 5 Detailed Scores (continued)

Attributes	Minimum	Maximum	Mean	StdDev
NRC-10-Expanded Positive	0	.877	.156	.27
SentiWordnet-posScore	0	2.586	.698	.827
SentiWordnet-negScore	-1.696	0	-.084	.173
Mpqa-posCount	0	1	.628	.483
Mpqa-negCount	0	1	.05	.219
BingLiu-negCount	0	1	.047	.211
AFINN-posScore	0	3	1.376	1.119
AFINN-negScore	-4	0	-.294	.593
S140-posScore	0	2.064	.334	.45
S140-negscore	-2.41	0	-.198	.374

The results represent the direct correlations between sentiment found in court cases that is paired to known negative sentiments used in the GPFLE process and content extracted from Tweets. Thus, one can hypothesize that negative sentiment associated with forms of fraud can be extracted from social media, as shown in the collective experiments' output.

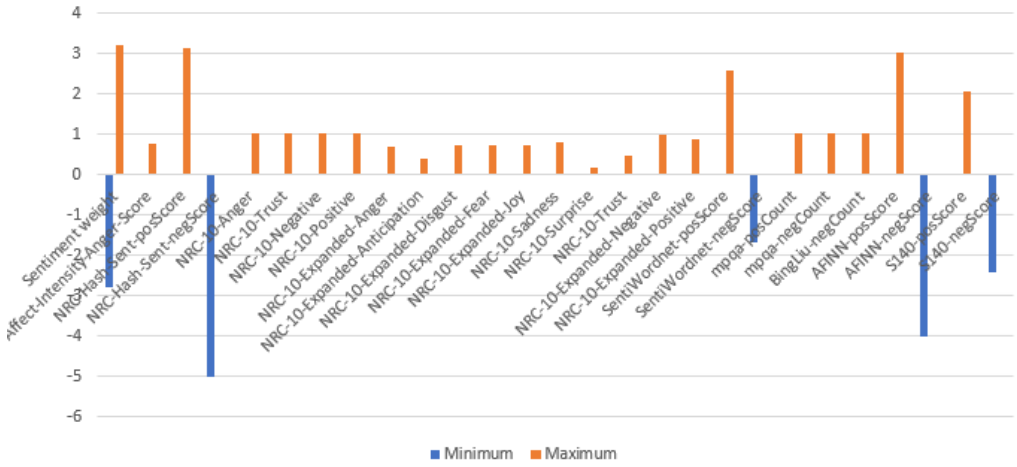


Figure 28. Artifact 5 Tweet negative emotion scores.

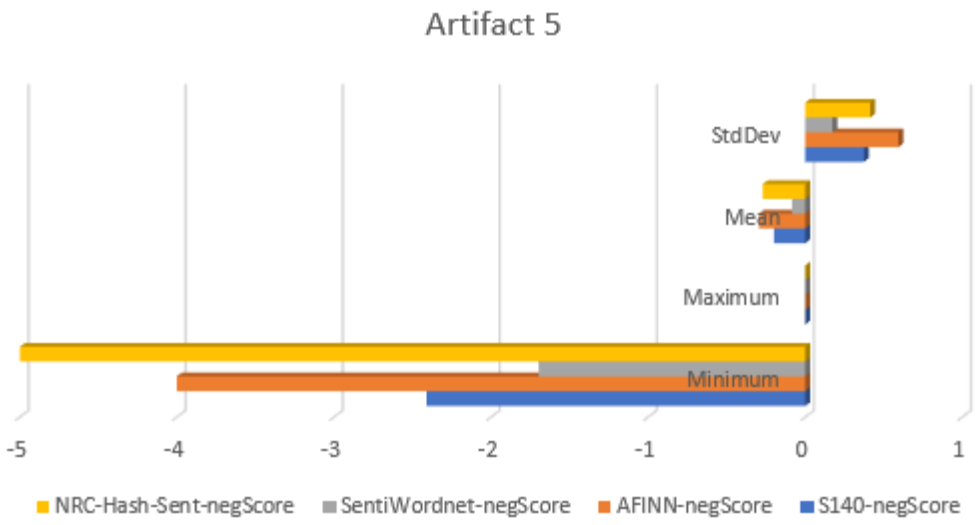


Figure 29. Artifact 5 negative summary.

The outcome classification scored compared slightly better than the previous experiment; Naïve Bayes (see Figures 30 –35) and correctly identified the classification instances 78% of the time. Like other experiments, the study continued to manually test various classifications and opted to use the Naïve Bayes classification.

Correctly Classified Instances	5864	78.0306 %
Incorrectly Classified Instances	1651	21.9694 %
Kappa statistic	0.7588	
Mean absolute error	0.0028	
Root mean squared error	0.0386	
Relative absolute error	23.5344 %	
Root relative squared error	50.3249 %	
Total Number of Instances	7515	

Figure 30. Artifact 5 Naïve Bayes classification – Part 1

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.000	0.001	0.000	0.000	0.000	-0.001	0.992	0.103	ability
0.333	0.001	0.182	0.333	0.235	0.245	0.997	0.146	accomplish
0.000	0.002	0.000	0.000	0.000	-0.002	0.997	0.242	active
0.000	0.002	0.000	0.000	0.000	-0.002	0.993	0.080	actively
0.500	0.001	0.421	0.500	0.457	0.458	0.998	0.416	admit
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	admitted
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	alone
0.000	0.002	0.000	0.000	0.000	-0.001	0.998	0.136	approval
0.333	0.001	0.222	0.333	0.267	0.271	0.999	0.290	authority
0.000	0.001	0.000	0.000	0.000	-0.000	0.998	0.113	avoided
0.615	0.001	0.615	0.615	0.615	0.615	0.999	0.761	award
0.000	0.001	0.000	0.000	0.000	-0.000	0.969	0.004	awarded
0.000	0.002	0.000	0.000	0.000	-0.002	0.995	0.103	benefit
0.250	0.003	0.091	0.250	0.133	0.149	0.997	0.154	benefits
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	best
0.277	0.024	0.337	0.277	0.304	0.278	0.961	0.358	better
0.000	0.001	0.000	0.000	0.000	-0.000	0.999	0.196	boosted
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.189	burden
0.000	0.000	0.000	0.000	0.000	-0.000	0.983	0.008	burdens
0.933	0.000	0.987	0.933	0.959	0.959	1.000	0.997	care
0.028	0.005	0.024	0.028	0.026	0.021	0.994	0.283	certain
0.000	0.001	0.000	0.000	0.000	-0.001	0.999	0.238	challenges
0.607	0.002	0.586	0.607	0.596	0.595	0.998	0.727	clean
0.000	0.000	0.000	0.000	0.000	-0.000	0.995	0.024	cleaner
0.578	0.002	0.650	0.578	0.612	0.611	0.998	0.721	clear
0.214	0.002	0.143	0.214	0.171	0.173	0.995	0.160	committed
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.119	complaint
0.727	0.003	0.286	0.727	0.410	0.455	0.998	0.309	confidence
0.167	0.001	0.125	0.167	0.143	0.144	0.998	0.174	confusion

Figure 31. Artifact 5 Naïve Bayes classification – Part 2.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.083	0.004	0.097	0.083	0.090	0.086	0.994	0.284	create
0.826	0.000	0.950	0.826	0.884	0.886	1.000	0.961	created
0.333	0.002	0.077	0.333	0.125	0.159	0.998	0.096	creates
0.000	0.001	0.000	0.000	0.000	-0.000	0.965	0.004	deceive
0.000	0.001	0.000	0.000	0.000	-0.000	0.976	0.006	deceived
0.000	0.000	0.000	0.000	0.000	-0.000	0.976	0.006	deception
0.500	0.001	0.500	0.500	0.500	0.499	0.999	0.688	defeat
0.250	0.002	0.333	0.250	0.286	0.287	0.998	0.411	defense
0.000	0.000	0.000	0.000	0.000	-0.001	0.999	0.323	delay
0.000	0.000	0.000	0.000	0.000	-0.000	0.969	0.004	determination
0.000	0.002	0.000	0.000	0.000	-0.001	0.995	0.053	determined
0.923	0.000	0.960	0.923	0.941	0.941	1.000	0.993	difficult
0.000	0.001	0.000	0.000	0.000	-0.000	0.997	0.073	dispute
0.000	0.001	0.000	0.000	0.000	-0.001	0.999	0.160	disregard
0.222	0.001	0.182	0.222	0.200	0.200	0.998	0.241	effective
0.000	0.000	0.000	0.000	0.000	-0.000	0.996	0.046	efficient
0.250	0.001	0.111	0.250	0.154	0.166	0.998	0.146	engaged
0.000	0.001	0.000	0.000	0.000	-0.001	0.995	0.064	engaging
0.000	0.001	0.000	0.000	0.000	-0.001	0.996	0.148	ensure
0.100	0.001	0.100	0.100	0.100	0.099	0.995	0.129	entitled
0.900	0.000	0.900	0.900	0.900	0.900	1.000	0.970	escape
0.000	0.000	0.000	0.000	0.000	-0.000	0.998	0.113	excluded
0.600	0.001	0.273	0.600	0.375	0.404	0.999	0.261	exclusive
0.000	0.000	?	0.000	?	?	0.976	0.005	exploiting
0.571	0.000	0.727	0.571	0.640	0.644	1.000	0.812	failed
0.375	0.001	0.429	0.375	0.400	0.400	0.999	0.387	failure
0.000	0.000	?	0.000	?	?	0.976	0.005	failures
0.286	0.004	0.229	0.286	0.254	0.252	0.993	0.228	fair
0.885	0.003	0.469	0.885	0.613	0.643	0.998	0.453	fit
0.917	0.000	1.000	0.917	0.957	0.957	1.000	1.000	fraud
0.704	0.003	0.842	0.704	0.767	0.766	0.998	0.913	free
0.800	0.001	0.444	0.800	0.571	0.596	0.999	0.426	friendly

Figure 32. Artifact 5 Naïve Bayes classification – Part 3.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.000	0.000	?	0.000	?	?	0.997	0.058	futile
0.750	0.001	0.875	0.750	0.808	0.808	0.998	0.893	giving
0.673	0.015	0.810	0.673	0.735	0.716	0.984	0.869	good
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	great
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.102	grossly
0.000	0.001	0.000	0.000	0.000	-0.000	0.999	0.226	harmed
0.000	0.001	0.000	0.000	0.000	-0.001	0.998	0.142	harsh
0.000	0.001	0.000	0.000	0.000	-0.000	0.999	0.156	hoax
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	illegal
0.000	0.000	0.000	0.000	0.000	-0.000	0.998	0.142	immoral
0.333	0.000	0.250	0.333	0.286	0.288	1.000	0.387	imposed
0.133	0.003	0.083	0.133	0.103	0.103	0.994	0.153	improve
0.000	0.002	0.000	0.000	0.000	-0.001	0.994	0.034	improvements
0.000	0.000	0.000	0.000	0.000	-0.000	0.965	0.004	inability
0.000	0.002	0.000	0.000	0.000	-0.002	0.996	0.124	increased
0.286	0.001	0.222	0.286	0.250	0.251	0.999	0.291	injured
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	injury
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.077	innovative
0.476	0.005	0.217	0.476	0.299	0.319	0.993	0.182	interest
0.400	0.001	0.333	0.400	0.364	0.365	0.999	0.327	interests
0.515	0.007	0.490	0.515	0.502	0.496	0.993	0.517	join
0.200	0.000	0.250	0.200	0.222	0.223	0.999	0.228	lawsuit
0.976	0.000	0.992	0.976	0.984	0.984	1.000	0.999	leave
0.481	0.002	0.481	0.481	0.481	0.480	0.998	0.506	legal
0.000	0.000	0.000	0.000	0.000	-0.000	0.998	0.102	liability
0.985	0.000	0.999	0.985	0.992	0.990	1.000	1.000	like
0.300	0.001	0.300	0.300	0.300	0.299	0.999	0.360	limited
0.000	0.000	0.000	0.000	0.000	-0.000	0.997	0.069	losses
0.809	0.000	0.950	0.809	0.874	0.876	1.000	0.967	low
0.609	0.003	0.412	0.609	0.491	0.499	0.997	0.409	lower
0.880	0.000	0.964	0.880	0.920	0.920	1.000	0.990	matter
0.000	0.000	0.000	0.000	0.000	-0.001	0.999	0.251	misleading

Figure 33. Artifact 5 Naïve Bayes classification – Part 4.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
1.000	0.000	0.923	1.000	0.960	0.961	1.000	1.000	negative
0.977	0.000	0.999	0.977	0.988	0.986	1.000	1.000	no
0.934	0.000	0.988	0.934	0.960	0.960	1.000	0.996	number
0.500	0.000	1.000	0.500	0.667	0.707	1.000	1.000	obstacles
0.000	0.000	0.000	0.000	0.000	-0.000	0.997	0.063	offend
0.133	0.002	0.222	0.133	0.167	0.170	0.996	0.330	opportunity
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.250	parties
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	pay
0.947	0.000	1.000	0.947	0.973	0.973	1.000	0.995	poor
0.750	0.001	0.450	0.750	0.563	0.580	0.999	0.438	prevent
0.920	0.000	0.988	0.920	0.952	0.952	1.000	0.993	problem
0.000	0.000	0.000	0.000	0.000	-0.001	0.995	0.105	profit
0.000	0.001	0.000	0.000	0.000	-0.001	0.996	0.081	profits
0.250	0.002	0.176	0.250	0.207	0.209	0.996	0.183	progress
0.000	0.002	0.000	0.000	0.000	-0.002	0.991	0.137	promise
0.154	0.002	0.100	0.154	0.121	0.122	0.996	0.160	promote
0.273	0.002	0.391	0.273	0.321	0.324	0.996	0.418	protect
0.000	0.000	0.000	0.000	0.000	-0.000	0.994	0.063	protects
0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.139	punish
0.000	0.000	?	0.000	?	?	0.997	0.069	reckless
0.273	0.001	0.231	0.273	0.250	0.250	0.998	0.245	refuse
0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.138	relief
0.869	0.000	0.964	0.869	0.914	0.914	1.000	0.975	respect
0.000	0.001	0.000	0.000	0.000	-0.000	0.963	0.004	respective
0.000	0.002	0.000	0.000	0.000	-0.002	0.992	0.100	responsible
0.545	0.001	0.429	0.545	0.480	0.483	0.999	0.426	risk
0.000	0.000	?	0.000	?	?	0.996	0.047	satisfied
0.000	0.001	0.000	0.000	0.000	-0.000	0.818	0.001	satisfy
0.424	0.006	0.448	0.424	0.436	0.429	0.993	0.418	share
0.000	0.002	0.000	0.000	0.000	-0.002	0.992	0.106	shared
0.514	0.002	0.559	0.514	0.535	0.533	0.997	0.622	sharing
0.500	0.001	0.250	0.500	0.333	0.353	0.999	0.262	significant

Figure 34. Artifact 5 Naïve Bayes classification – Part 5.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
	0.000	0.000	?	0.000	?	?	0.965	0.004	sluggish
	0.000	0.000	?	0.000	?	?	0.998	0.067	sophisticated
	0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.196	stinky
	0.714	0.000	0.870	0.714	0.784	0.787	1.000	0.919	stopped
	0.000	0.000	0.000	0.000	0.000	-0.000	0.978	0.006	substantial
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	successful
	0.778	0.000	0.875	0.778	0.824	0.825	1.000	0.940	suffer
	0.667	0.000	0.667	0.667	0.667	0.666	1.000	0.768	suffered
	0.375	0.001	0.333	0.375	0.353	0.353	0.999	0.378	superior
	0.510	0.030	0.329	0.510	0.400	0.389	0.970	0.315	thanks
	0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.125	touted
	0.552	0.005	0.481	0.552	0.514	0.510	0.995	0.457	trust
	0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.149	trusted
	0.000	0.001	0.000	0.000	0.000	-0.000	0.843	0.001	uncertain
	0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.139	unethical
	0.000	0.000	0.000	0.000	0.000	-0.000	0.999	0.226	unfair
	0.158	0.002	0.200	0.158	0.176	0.176	0.996	0.222	united
	0.000	0.000	0.000	0.000	0.000	-0.000	0.983	0.008	unjust
	0.000	0.002	0.000	0.000	0.000	-0.002	0.992	0.162	value
	0.286	0.001	0.200	0.286	0.235	0.238	0.998	0.179	values
	0.000	0.000	0.000	0.000	0.000	-0.001	0.999	0.323	victims
	0.333	0.000	0.250	0.333	0.286	0.288	0.999	0.294	violate
	0.000	0.000	0.000	0.000	0.000	-0.000	0.980	0.007	violated
	0.000	0.000	0.000	0.000	0.000	-0.000	0.987	0.011	violation
	0.000	0.000	0.000	0.000	0.000	-0.000	0.980	0.007	violations
	0.000	0.001	0.000	0.000	0.000	-0.001	0.997	0.084	virtue
	0.000	0.001	0.000	0.000	0.000	-0.000	0.999	0.196	vision
	0.900	0.001	0.980	0.900	0.938	0.937	0.999	0.987	well
	0.833	0.000	0.959	0.833	0.892	0.893	1.000	0.978	worth
Weighted Avg.	0.780	0.004	?	0.780	?	?	0.995	0.828	

Figure 35. Artifact 5 Naïve Bayes classification – Part 6.

Within the context of this experiment, the TPR's weighted average came to *0.780* and considered acceptable when comparing a perfect rate of 1.0. In contrast, the FPR's weighted average scored a *0.004*; comparable to the previous experiment. The weighted average with the Precision was undetermined through the automated process within WEKA and manual calculation with TPR / (TPR + FPR) produced an average of 0.995; acceptable, but the appears to be slightly higher than previous experiments, but comparable to experiment 3.

Like other experiments, this experiment produced TPR and FPR numbers without values, and could represent no corresponding class assignments. For instance, zero-valued classes accounted for thirty-six negative lexicons; *avoided, burdens, complaint, deceive, deceived, deception, delay, dispute, disregard, excluded, exploiting, failures, futile, grossly, harmed, harsh, hoax, immoral, inability, liability, losses, misleading, offend, punish, reckless, sluggish, touted, uncertain, unethical, unfair, unjust, victims, violated, violation,* and violations. Class values greater than zero and less than one accounted for eight negative lexicons; *burden, confusion defeat, delay, difficult, failed, failure,* and *fraud*. And the last class with values equal to one accounted six negative lexicons; *alone, illegal, injury, negative,* and *pay* (see Appendix U). Also, this experiment's classification of all TPR classes might be identifiers with lexicons associated with fraud and differ from the preceding experiments.

Although this experiment supplied more data with 7,515 class instances, the outcome provided an explanation as to specific attributes set to undefined. The initial assessment postulated the lack of data caused similar results in previous experiments. In this instance, that is not the case. Having more data did not prove to be more beneficial,

and it is believed the disparity is a result from the type of data being processed, same overall court system, but a uniqueness in content. Further examination has shown the weighted averages for Precision, F-Measure, and MCC was uncalculated; some data could not be assigned to a given class.

For these attributes, the following outcome was possible. First, the weighted averages for Recall yielded 0.780 and like the previous experiment considered good in terms of correctly labeling lexicons. Second, ROC yielded 0.995 as its optimal threshold of false positives. The supportive outcome provided the identification of some lexicon classes conducive in vetting social media tweets and could help mitigate insider threats from the most earliest onset.

The theories that correlate best with this case are protection motivation theory (PMT) and routine activity theory (RAT). Protection motivation theory represents the cognitive processes to mediate the persuasive effects of a fear appeal by arousing protection motivation. In this case, it appears dwindling sales was a motivation and according to Maddux and Rogers (1983), the protection motivation came from self-preservation with keeping the business afloat. Furthermore, one could theorize the danger felt by the manufacturer might be construed with the fear from competitors and led to the business finding suitable targets, the consumer and in alignment with RAT.

Lab Experiment 6: Artifact 6

Experiment 6 was identified by Artifact ID 3-16-CV-01547. In this case, an employee of a business exhibited fraudulent practices. As in the previous instance, the study leveraged an all-encompassing and differing angle for fraud through an employee associated with the banking industry. The experiment included an increase of classifications over the previous experiment, yet similarities was discovered through court data.

Data were retrieved from The United States District Court, Southern District of California. The Southern District provided data from a bank that demonstrated a former contractor's unwillingness to return the bank's laptop and its proprietary software. In addition to the contractor's possession of the laptop came threats to sell the bank's private and sensitive information to anyone willing to pay the highest price. According to USCourts.gov (2020) Southern District of California:

> The Bank seeks immediate injunctive relief from this Court to compel Deaver to refrain from disclosing and selling any of the Bank's trade secret information and to order him to return to the Bank the property that he literally stole. The Bank will be irreparably harmed if Deaver is not ordered to return the laptop and source

> code that he stole and claims to have provided and disclosed to third parties. If there is a case worthy of a temporary restraining order to stop the criminal conduct of a rogue contractor who is intent on carrying out his criminal threats, this is it. (p. 2)

Just as earlier experiments required a data conversion from PDF using OCR, the same was applied to this experiment. The documents appeared to be scanned from within the courts and into an electronic format, with many pages not correctly aligned during the scanning process and required the use of OCR software.

This dissertation study followed the GPFLE process for Lab Experiment 6 and supplemented the process with the successive activities in capturing all sentiment attributes (see Table 8) with the lowest negative sentiment scores from the most prominent scores (see Figure 36). An alternate summary view (see Figure 37) is also presented and shows the lowest four negative sentiment scores and validation using the DMNB classification (see Figures 38-43). A common theme between experiments was the four negative (see Appendix K) sentiment scores. For example, the S140-negScore (-2.148), AFINN-negScore (-4.0), SentiWordnet-negScore (-2.148), and NRC-Hash-Sent-negScore (-4.999).

Table 8
Artifact 6 Detailed Scores

Attributes	Minimum	Maximum	Mean	StdDev
Sentiment weight	-3.2	3.1	.712	1.542
Hits	0	239	71.032	82.094
NRC-Affect-Intensity-Anger-Score	0	.882	.033	.133
NRC-Hash-Sent-posScore	0	5	.025	.482
NRC-Hash-Sent-negScore	-4.999	0	-.382	.454
NRC-10-Anger	0	1	.064	.245
NRC-10-Trust	0	1	.106	.307
NRC-10-Negative	0	1	.122	.327
NRC-10-Positive	0	1	.209	.407

Table 8
Artifact 6 Detailed Scores (continued)

Attributes	Minimum	Maximum	Mean	StdDev
NRC-10-Expanded-Anger	0	.713	.026	.069
NRC-10-Expanded-Anticipation	0	.314	.063	.089
NRC-10-Expanded-Disgust	0	.464	.012	.035
NRC-10-Expanded-Fear	0	.74	.24	.08
NRC-10-Expanded-Joy	0	.724	.108	.217
NRC-10-Sadness	0	.807	.32	.119
NRC-10-Surprise	0	.149	.025	.043
NRC-10-Trust	0	.684	.118	.18
NRC-10-Expanded-Negative	0	.956	.076	.169
NRC-10-Expanded Positive	0	.883	.207	.304
NRC-10-SentiWordnet-posScore	0	1.539	.335	.476
NRC-10-SentiWordnet-negScore	-1.067	0	-.094	.221
Mpqa-posCount	0	1	.512	.5
Mpqa-negCount	0	1	.147	.354
BingLiu-negCount	0	1	.151	.358
AFINN-posScore	0	3	.795	.99
AFINN-negScore	-4	0	-.477	.938
S140-posScore	0	1.707	.32	.381
S140-negscore	-2.148	0	-.194	.407

The results represent the direct correlations between sentiment found in court cases that is paired to known negative sentiments used in the GPFLE process and content extracted from Tweets. Thus, one can hypothesize that negative sentiment associated with forms of fraud exists within social media.

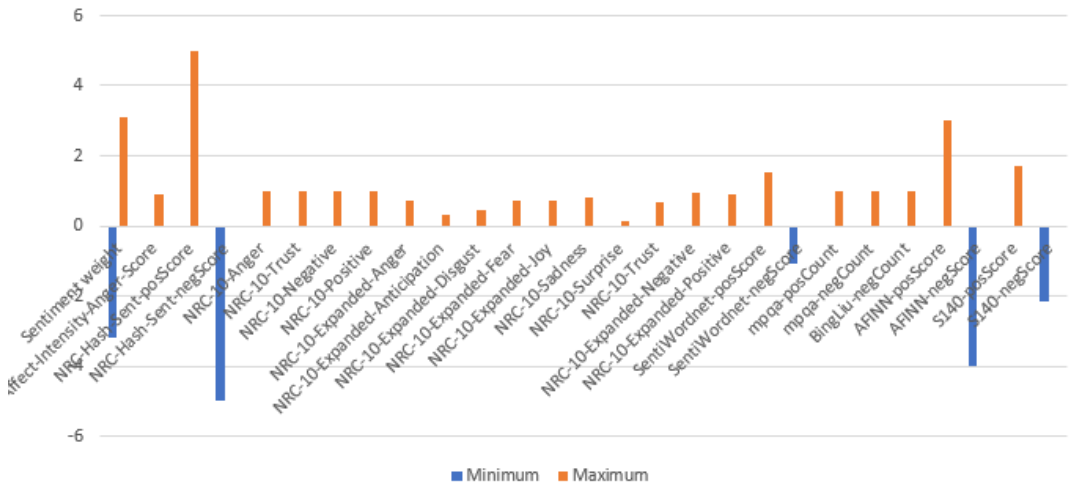


Figure 36. Artifact 6 Tweet negative emotion scores.

The classification scored comparatively better than the previous experiment, correctly identified the classification instances 92% of the time, and yielded a Kappa score of 0.9257. The important takeaway from these results is the overall average for the classifications of all experiments rather than this set of results.

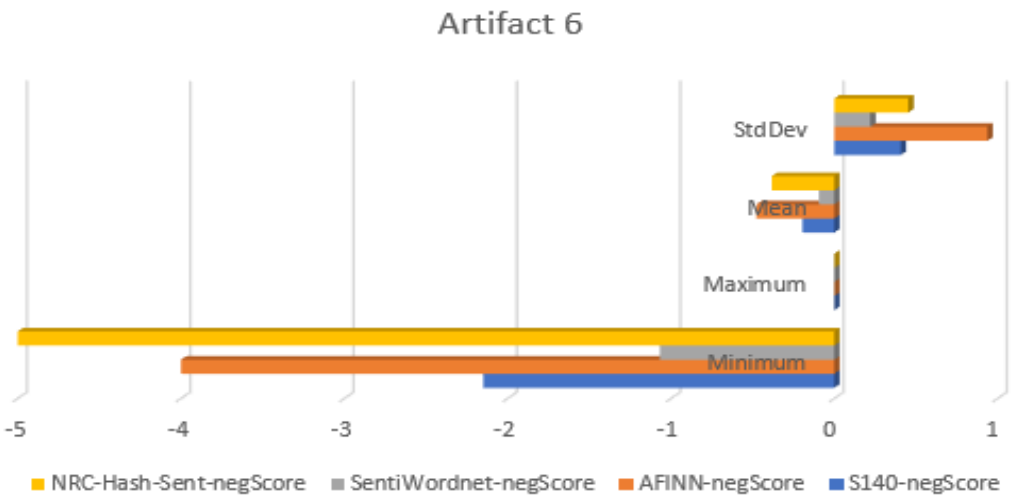


Figure 37. Artifact 6 negative summary.

Correctly Classified Instances	3037	92.903 %
Incorrectly Classified Instances	232	7.097 %
Kappa statistic	0.9257	
Mean absolute error	0.0052	
Root mean squared error	0.0392	
Relative absolute error	33.9152 %	
Root relative squared error	45.0434 %	
Total Number of Instances	3269	

Figure 38. Artifact 6 Discriminative Multinomial Naïve Bayes classification – Part 1.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.000	0.000	?	0.000	?	?	0.997	0.429	abuse
0.000	0.000	?	0.000	?	?	0.683	0.001	admits
0.000	0.000	?	0.000	?	?	0.989	0.234	admitted
0.250	0.001	0.333	0.250	0.286	0.287	0.998	0.493	advantage
1.000	0.005	0.884	1.000	0.938	0.938	1.000	1.000	agree
1.000	0.000	0.952	1.000	0.976	0.976	1.000	1.000	agreed
1.000	0.004	0.536	1.000	0.698	0.730	1.000	1.000	agreement
1.000	0.000	0.982	1.000	0.991	0.991	1.000	1.000	alone
0.000	0.000	?	0.000	?	?	0.816	0.004	asset
0.000	0.000	?	0.000	?	?	0.827	0.002	assurances
0.167	0.000	1.000	0.167	0.286	0.408	1.000	1.000	authority
0.000	0.000	?	0.000	?	?	0.998	0.167	avoided
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	award
0.250	0.002	0.286	0.250	0.267	0.266	0.998	0.443	benefit
0.875	0.000	1.000	0.875	0.933	0.935	1.000	1.000	benefits
1.000	0.005	0.706	1.000	0.828	0.838	1.000	1.000	certain
1.000	0.001	0.938	1.000	0.968	0.968	1.000	1.000	clear
1.000	0.000	0.933	1.000	0.966	0.966	1.000	1.000	committed
0.000	0.000	?	0.000	?	?	0.992	0.353	committing
0.000	0.000	?	0.000	?	?	0.665	0.001	compelled
0.000	0.000	?	0.000	?	?	0.817	0.002	competitive
0.000	0.000	?	0.000	?	?	0.980	0.031	complained
0.000	0.000	?	0.000	?	?	0.983	0.062	complaint
0.000	0.000	?	0.000	?	?	0.989	0.101	consent
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	created
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	credit
1.000	0.001	0.750	1.000	0.857	0.865	1.000	1.000	crime
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	criminal
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	damage

Figure 39. Artifact 6 Discriminative Multinomial Naïve Bayes classification – Part 2.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.000	0.000	?	0.000	?	?	0.997	0.468	damaging
1.000	0.000	0.967	1.000	0.983	0.983	1.000	1.000	dear
1.000	0.001	0.714	1.000	0.833	0.845	1.000	1.000	demand
0.000	0.000	?	0.000	?	?	0.666	0.001	demanded
0.000	0.000	?	0.000	?	?	0.999	0.750	demanding
1.000	0.002	0.722	1.000	0.839	0.849	1.000	1.000	destroy
0.833	0.000	1.000	0.833	0.909	0.913	1.000	1.000	destroying
0.000	0.000	?	0.000	?	?	0.992	0.156	destruction
0.000	0.000	?	0.000	?	?	0.991	0.125	determined
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	difficult
0.000	0.000	?	0.000	?	?	0.955	0.014	disregard
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	easily
0.000	0.000	?	0.000	?	?	0.992	0.121	efficient
0.000	0.000	?	0.000	?	?	0.999	0.750	engaged
0.000	0.000	?	0.000	?	?	0.989	0.197	engagement
0.950	0.000	1.000	0.950	0.974	0.975	1.000	1.000	enjoyed
0.000	0.000	?	0.000	?	?	0.976	0.025	ensuring
0.000	0.000	0.000	0.000	0.000	-0.001	0.995	0.383	entitled
1.000	0.001	0.909	1.000	0.952	0.953	1.000	1.000	excuse
0.000	0.000	?	0.000	?	?	0.464	0.001	extends
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	failed
0.500	0.000	1.000	0.500	0.667	0.707	1.000	1.000	failing
0.625	0.000	1.000	0.625	0.769	0.790	1.000	1.000	failure
0.000	0.000	?	0.000	?	?	1.000	1.000	favor
0.000	0.000	?	0.000	?	?	0.978	0.023	favors
1.000	0.002	0.632	1.000	0.774	0.794	1.000	0.849	fraud
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	giving
1.000	0.000	0.997	1.000	0.998	0.998	1.000	1.000	great
0.222	0.001	0.500	0.222	0.308	0.332	1.000	0.980	greater
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	gross
0.000	0.000	?	0.000	?	?	0.897	0.014	hacked
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	harm
0.000	0.000	?	0.000	?	?	0.900	0.006	harmed
1.000	0.004	0.721	1.000	0.838	0.848	1.000	1.000	illegal

Figure 40. Artifact 6 Discriminative Multinomial Naïve Bayes classification – Part 3.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
0.000	0.000	?	0.000	?	?	0.679	0.001	improvements
0.000	0.000	?	0.000	?	?	0.994	0.165	integrity
0.000	0.000	?	0.000	?	?	0.854	0.003	intellectual
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	interest
0.400	0.000	1.000	0.400	0.571	0.632	1.000	1.000	lawsuit
1.000	0.004	0.659	1.000	0.794	0.810	1.000	0.999	legal
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	lies
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	limited
1.000	0.001	0.886	1.000	0.939	0.941	1.000	1.000	loss
1.000	0.003	0.921	1.000	0.959	0.958	1.000	1.000	lost
1.000	0.006	0.701	1.000	0.825	0.835	1.000	1.000	low
1.000	0.002	0.948	1.000	0.974	0.973	1.000	1.000	matter
0.000	0.000	?	0.000	?	?	0.507	0.001	merits
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	number
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	original
1.000	0.001	0.850	1.000	0.919	0.922	1.000	0.987	parties
1.000	0.001	0.966	1.000	0.983	0.983	1.000	1.000	party
1.000	0.005	0.952	1.000	0.976	0.973	1.000	1.000	please
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	prevent
0.000	0.000	?	0.000	?	?	0.935	0.007	preventing
0.000	0.000	?	0.000	?	?	1.000	1.000	profit
0.000	0.000	?	0.000	?	?	0.987	0.079	profits
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	promise
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	promised
0.000	0.000	?	0.000	?	?	0.888	0.006	promises
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	protect
0.333	0.000	1.000	0.333	0.500	0.577	1.000	1.000	protected
0.000	0.000	0.000	0.000	0.000	-0.001	1.000	0.903	refused
0.000	0.000	?	0.000	?	?	0.896	0.003	refusing
0.667	0.000	1.000	0.667	0.800	0.816	1.000	1.000	relief
1.000	0.004	0.824	1.000	0.904	0.906	1.000	1.000	respect
0.000	0.000	?	0.000	?	?	0.353	0.001	respectfully
0.000	0.000	?	0.000	?	?	0.794	0.002	restricting

Figure 41. Artifact 6 Discriminative Multinomial Naïve Bayes classification – Part 4.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Lexicon
	0.857	0.000	0.923	0.857	0.889	0.889	1.000	0.938	secure
	0.000	0.000	?	0.000	?	?	0.956	0.021	secured
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	security
	1.000	0.004	0.885	1.000	0.939	0.939	1.000	1.000	share
	1.000	0.000	0.923	1.000	0.960	0.961	1.000	1.000	steal
	0.778	0.000	1.000	0.778	0.875	0.882	1.000	0.977	stolen
	1.000	0.000	0.996	1.000	0.998	0.998	1.000	1.000	stop
	0.000	0.000	?	0.000	?	?	0.257	0.000	substantial
	1.000	0.001	0.917	1.000	0.957	0.957	1.000	1.000	success
	1.000	0.000	0.900	1.000	0.947	0.949	1.000	1.000	suffer
	0.167	0.000	1.000	0.167	0.286	0.408	0.997	0.368	suffered
	1.000	0.004	0.923	1.000	0.960	0.959	1.000	1.000	support
	0.000	0.000	?	0.000	?	?	0.976	0.020	suspected
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	threat
	0.000	0.000	?	0.000	?	?	0.995	0.416	threatened
	0.600	0.000	1.000	0.600	0.750	0.774	1.000	0.927	threatening
	0.000	0.000	?	0.000	?	?	0.798	0.002	threatens
	0.000	0.002	0.000	0.000	0.000	-0.002	0.997	0.277	threats
	0.000	0.000	?	0.000	?	?	0.999	0.369	unacceptable
	0.000	0.000	?	0.000	?	?	0.905	0.005	unethical
	0.526	0.001	0.833	0.526	0.645	0.661	1.000	1.000	united
	0.000	0.000	?	0.000	?	?	0.906	0.003	unjust
	0.833	0.000	1.000	0.833	0.909	0.913	1.000	1.000	valuable
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	value
	0.000	0.000	?	0.000	?	?	0.978	0.027	violate
	0.000	0.000	?	0.000	?	?	0.808	0.002	violated
	0.000	0.000	?	0.000	?	?	0.938	0.005	violation
	0.000	0.000	?	0.000	?	?	1.000	0.589	virtue
	1.000	0.003	0.973	1.000	0.987	0.985	1.000	1.000	well
	0.385	0.000	0.833	0.385	0.526	0.565	1.000	0.984	worthy
Weighted Avg.	0.929	0.002	?	0.929	?	?	0.997	0.960	

Figure 42. Artifact 6 Discriminative Multinomial Naïve Bayes classification – Part 5.

Within the context of this experiment, the TPR's weighted average came to *0.929* and considered acceptable when comparing a perfect rate of 1.0. In contrast, the FPR's weighted average scored a *0.002*; comparable to other experiments. Like the previous experiment, the weighted average with Precision was undetermined through the automated process within WEKA and manual calculation with TPR / (TPR + FPR) produced an average of 0.492; although this might be acceptable, the value is definitely lower than previous experiments.

As in other experiments, this experiment produced TPR and FPR numbers without values, and represents instances where the data cannot be assigned to a class. Instances of zero-valued classes accounted for twenty-four negative lexicons; *abuse, avoided, complained, complaint, demanded, demanding, destruction, disregard, hacked, harmed, preventing, refused, refusing restricting, suspected, threatened, threatens,*

threats, unacceptable, unethical, unjust, violate, violated, and *violation*. Class values greater than zero and less than one accounted for seven negative lexicons; *destroying, failing, failure, lawsuit, steal, suffered,* and *threatening*. The last class with values equal to one accounted for twelve negative lexicons; *alone, crime, criminal, damage, demand, destroy, difficult, failed, fraud, gross, harm,* and *illegal* (see Appendix V). Also, this experiment's classification of all TPR classes might be identifiers with lexicons associated with fraud and differ from the preceding experiments.

Analogous to the previous experiment, similarities exist in the failure to compute the weighted averages for the Precision, F-Measure, and MCC; all attributed to some of the data. Initially, it was believed to be caused by the lack of data, but in this instance, it appears to be related to instances not being able to have class assignments. Additionally, the weighted average for Recall at 0.929 and ROC of 0.997; an optimal threshold.

Correspondingly to other experiments interpretation with behavioral theories, RAT best applies to this case. In this instance, the contractor appeared to knowingly select a target thought to be incapable of defending itself, which revealed the absence of capable guardians against crime (Cohen & Felson, 1979). However, what the contractor did not realize at the time was the bank's trade secrets on the laptop and the willingness to prosecute based on theft.

Findings

Based on known and well-established negative sentiments, the study revealed that the fraudulent sentiments that exist in the legal system also exist in social media. This dissertation study applied an identical methodology in each lab experiment and found cynical sentiments within the publicly available Twitter tweets, correlation to unique and negative social media lexicons with attributes from a fraudulent context, and documented within the legal system from cases involving some degree of fraud. While some lab experiments either experienced a lack of data or data not assigned with the listed classes, many instances held values supportive to the outcome (see Appendix Q, R, S, T, U, and V).

The upfront notation to use Auto-WEKA to select the best classification algorithm was not the best option and was changed due to precisions of correctly identifying instances within the experiments. The experiments used three classifications: Random Forest, Naïve Bayes, and DMNB. Although this dissertation study experienced technical challenges earlier in the research; for instance, an essential WEKA plug-in kept causing problems, yet the problems were overcome. Other time-intensive operations appeared unique because the research operated within one physical location and did not leverage cloud-based technologies.

Each experiment's ending required a case review from each set of court documents. These case reviews were paired to one or more behavioral theories and found six cases that aligned with the RAT, two cases that aligned with the TPB, one case that aligned with the TRA, and one case that aligned with the protection motivation theory. Not all anticipated theories were applicable, yet this dissertation study found one

behavioral theory to be more prevalent in many lab experiments. RAT appeared to touch all aspects of court cases. This study's findings confirmed Cohen and Felson's (1979) behavioral theory was demonstrated through criminal acts coming down to a merging of offenders and targets, the element of timing, and not having any forms of protection against criminal activities become relevant predictors.

Summary of Results

Sentiment scoring to include relevant behavioral theories, classification variables, and exercise specific data (see Table 9) is outlined. All experiments shared three core commonalities; all relatable to RAT, used Hutto and Gilbert's (2014) 7,063 sentiment lexicons, and all accessed the same 20,000 Twitter tweets. Experiments 2-6 reached the lower negative NRC-Hash-SentnegScore of -4.990; nearly the lowest possible value of -5.0. Within the context of negativity, this represents tweets of sadness, anger, fear, or disgust and is outlined in Bravo-Marquez et al.'s (2015) works (see Appendix K).

Experiments 3,5, and 6 produced the lowest scoring for SentWordnet-negScore, and according to Baccianella et al. (2010), the correlation related to the top-ranked negative synsets. Similarly, experiments 1,3,4,5, and 6 showed a fair amount of negativity in the AFINN-negScore with a low value of -4.0. Furthermore, the S140-negScore for experiments 3-6 appeared to demonstrate negativity in emotion-aware tweets.

Also, all experiments included various classifications, including Trees Random Forest, Naïve Bayes, and Discriminative Multinomial Naïve Bayes. Kappa values within each experiment provided greater accuracy and according to Sahoo's (2013) findings, finer accuracy is achieved when values are higher than zero. Experiments 1,2,3, and 6

showed nearly perfect agreements with Kappa (see Appendix M) and experiments 4 and 5 demonstrated substantial agreements.

Examining each experiment's TPR, provided acceptable values and representative of positive classes correctly classified by a model is achieved (Azar et al., 2014) with experiments 1-6 yielding the following respective values of 0.873, 0.850, 0.948, 0.700, 0.780, and 0.929. In contrast, the FPR represents the fraction of negative classes that are identified as positive, and in these instances, all experiments appeared to score exceptionally low. All six experiments yielded the following respective values of 0.039, 0.009, 0.003, 0.005, 0.004, and 0.002. Ideally, the class precision's weighted averages would have been closer to 1.0 and represent the exactness of a classifier (Kaur & Saini, 2015) and correlate to fewer false positives. In half the experiments, 1,4, and 6, the FPR came to 0.539, 0.736, and 0.492. Conversely, experiments 2,3, and 5 yielded 0.999, 0.958, and 0.995.

The uniqueness between experiments is directly contributed to the content within the obtained legal documents. While each court case had anywhere between 1 and 36 pages of transcripts, it appears each showed relevance in the number of collected instances processed by the classifiers. For example, each respective experiment yielded 157, 206, 1,846, 5,926, 7,515, and 3,269 records of data. The overall holistic capture of data appears to provide a glimpse into the value of traversing tweets to identify possible insider threats. Furthermore, while limited in scope to fraudulent cases, there appears to be relevancy when examining data within the legal system and finding likenesses within publicly available tweets. And lastly, organizations could leverage these findings within the preemployment vetting of future employees and business associates.

Table 9
Summary of Results

Attribute	Description	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6
NRC-Hash-Sent-negScore	Sadness, anger, fear, and disgust	-1.244	-4.990	-4.990	-4.990	-4.990	-4.990
SentiWordnet-negScore	Top ranked negative synsets	-0.344	-0.291	-1.067	-0.208	-1.696	-1.067
AFINN-negScore	Slang, obscene words, web jargon	-3.000	-1.174	-4.000	-4.000	-4.000	-4.000
S140-negScore	Emotion aware negative tweets	-0.180	-0.180	-2.148	-4.990	-2.410	-2.148
Theory	Applicable theory	RAT	RAT/TPB	RAT/TPB	RAT/TRA	RAT/PMT	RAT
Instances	Records of data	157	206	1,846	5,926	7,515	3,269
Classification	Applied classifier	Trees random forest	Trees random forest	Discriminative multinomial	Naïve Bayes	Naïve Bayes	Discriminative multinomial
Kappa	Kappa score	0.8055	0.8106	0.9449	0.6822	0.7588	0.9257
TPR	True positive rate	0.873	0.850	0.948	0.700	0.780	0.929
FPR	False positive rate	0.039	0.009	0.003	0.005	0.004	0.002
Precision	Precision	0.539	0.999	0.958	0.736	0.995	0.492
Negative sentiment lexicons	Maximum available	7,063	7,063	7,063	7,063	7,063	7,063
Twitter tweets	Maximum available	20,000	20,000	20,000	20,000	20,000	20,000
Legal documents	Pages	1	22	23	36	33	23
Negative Classes	Lexicons	2	4	30	58	48	43

Chapter 5

Conclusions, Implications, Recommendations, and Summary

Conclusions

This dissertation study found correlations from data extracted from both fraud and social media inputs. The benefits of machine learning, in-depth analysis of data, and coupling of theories that elaborated on behavioral aspects with insider threat mindsets (see Figure 43) produced evidence that proved beneficial in mitigating an insider threat at an early onset.

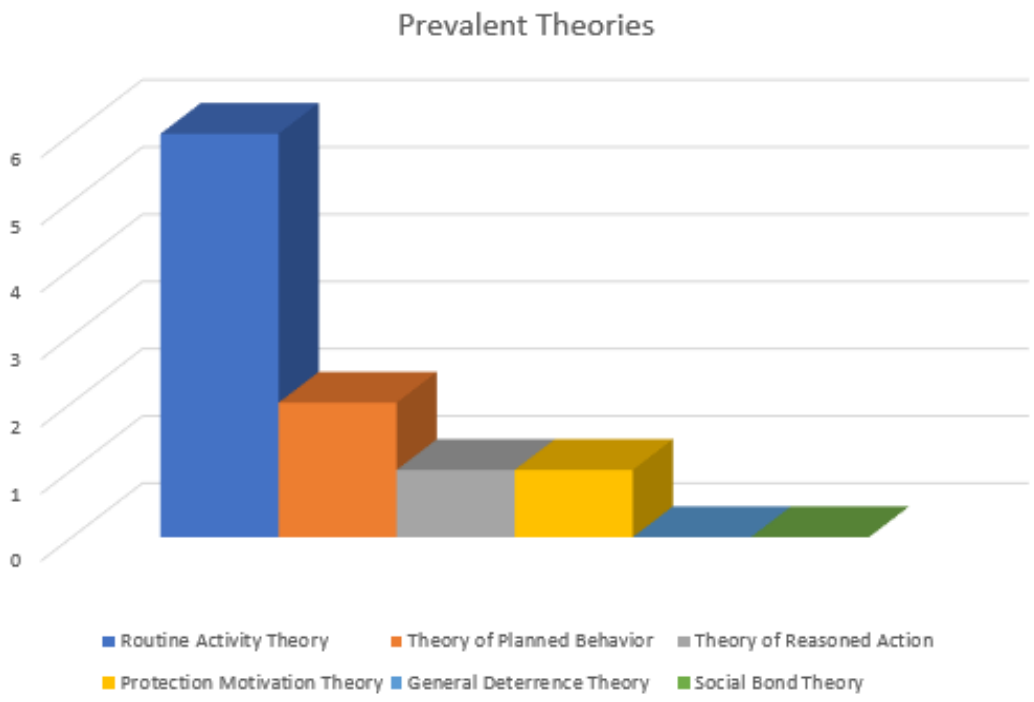


Figure 43. Prevalent theories.

Conversely, this dissertation study aimed to gather evidence to meet the objectives created by the IT artifacts in proving useful in threat predictions (see Figure 44). Three objectives; first, aimed to draw on correlations between negative sentiments found in various fraudulent cases and the same lexicons found in social media tweets. The goal within this context was to provide a preemployment vetting tool to identify individuals showing similarities with insider threat behaviors. Second, aimed to go beyond natural language processing's shortcomings to find commonalities fraudulent contexts within all lab experiments, including behavioral theories common and paramount among all fraudulent cases. And third, the analysis of relevant data attributed with varying degrees of fraud.

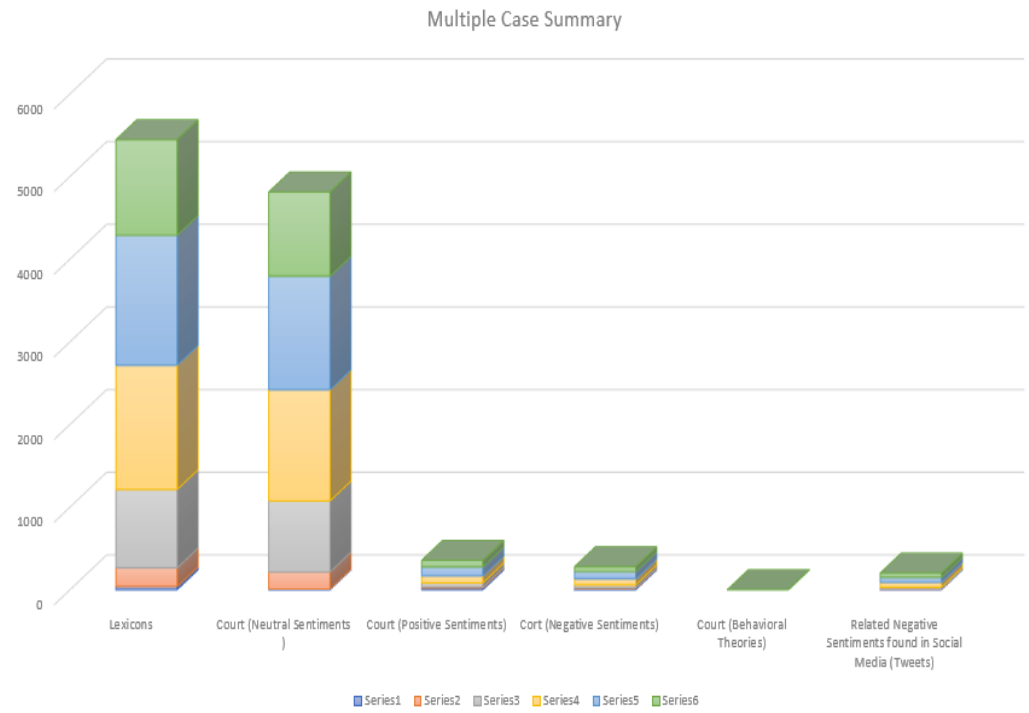


Figure 44. Multiple case summary.

This dissertation study contributed to positive outcomes; however, limitations surfaced. First, focus was exclusively on English content from fraudulent cases and collected social inputs from Tweets. A more elaborate study could build upon threats in other languages and pull data from Instagram, YouTube, or similar platforms. Second, fraud-related insider threats only scratch the surface with a few types of threats. Other forms of insider threats could include other forms of white-collar crimes. According to H.G. Legal Resources (2020), other forms of crimes include the following:

> Insurance fraud including automotive insurance, homeowner insurance, medical insurance and Medicaid insurance may all involve people who attempt to commit fraud. Insider trading, securities fraud, hedge fund fraud and stock manipulation are white collar crimes that may be committed when stocks or bonds are involved. Computer fraud, wire fraud and mail fraud may also be the result of white-collar crimes. White collar crimes also include identity theft, mortgage broker fraud, bribery, embezzlement, and tax evasion. These crimes are some of the most common types of white-collar crimes. These crimes all have the similarity that the ultimate goal is to receive some type of economic or financial gain. (para. 4)

Third, the initial assessment gravitated in using Facebook as a significant source of inputs; however, during testing with sample data, the source did not produce sufficient data for analysis. At that time, Twitter was selected and required additional time for approvals. This dissertation study applied for developer's access and waited weeks for approval, which delayed the research process. In hindsight, the approval process should have been initiated earlier in the research process. Fourth, the lab experiments provided insight into answering the research questions.

Research Question 1 asked: Was there sufficient literature on insider threat mitigation strategies? The answer to this question was yes. This dissertation study gathered a fair amount of literature that was published between 2013 and 2019. The literature covered three core areas examined by the study and included monitoring and profiling, rulemaking and policies, and employment vetting.

Under the umbrella of monitoring and profiling, Cole (2015) pointed out the majority of current mitigation efforts include monitoring of both internal and external networks. In this instance, Cappelli et al. (2009) found the level of practicality did not appear the best solution for an institution, while Kühn et al. (2017) found any benefits with monitoring events not fruitful when dealing with the analysis from intrusion-detection monitoring. Supportive to a monitoring analysis, Benferhat et al. (2013) suggested when dealing with a dynamic and changing environment, referencing a baseline can become an ineffective approach. In other instances, Shaw (2006) discovered supervisors required the knowledge to know whether or not an employee is disgruntled through evaluating risk factors and became apparent there is a delay within the mitigation strategy. Comparatively, Hubballi and Suryanarayanan (2014) delved deep into missing key alerts through false negatives within SIEM technologies. Furthermore, Vilendečić et al. (2017) suggested key implementations within SIEM required lowering false-positives or preventive action cannot be taken at the right time (Ambre & Shekokar, 2015).

As the literature suggests, policy enforcement does very little in terms of changing employees' mindsets. For instance, literature has demonstrated Acceptable Use Policies are fairly common, yet becomes worthless if employees do not become aware of them (Alshboul & Streff, 2017) and Gallagher et al. (2015) suggested an inadequate implementation will not alter users' postures towards insider threat preventions. Linkov et al. (2019) theorized policies that are over and under-regulated can become exploitable. Antoniou et al. (1999) suggested competing principles can drive conflicts, then promote an unclear direction to employees who rely on voluntary compliance and cooperation (Pelton, 2017). Moreover, according to Bauer (2017), despite having policies in place at

an organizational level, employees intentionally are noncompliant; due to carelessness, poor knowledge, or clear intention to act dishonestly (Nawawi & Salin, 2018).

The literature on employee vetting is the basis for this study and some approaches were demonstrated within failures within current vetting practices. According to Kühn and Nieman (2017), flaws can be contributed to the over-reliance of information collected from the employee, and regardless of the layers of vetting, people still pass through the process. Lomas (2019) discovered personnel responsible for vetting officials is flawed, along with Jeske et al.'s (2019) findings of faults with voluntary employee disclosures; not solidify sound practices (Hielscher & Waghorn, 2015). Lastly, Roulin and Bourdage (2017) discovered it is possible to uncover personality traits during the interviewing process, while Maasberg et al. (2015) postulated negative attitudes, triggers, motives, malicious intent, and motives are security concerns that need to be addressed. All reviewed literature directly supports insider threat mitigation attempts from the past and all relevant to the outcome of this study.

In summary, non-behavioral literature reviews spanned approximately 400 articles and selected the top 55 to cover 1999, through 2020, while behavioral literature included six theories dating back to 1979. All literature encompassed a multitude of content directly related to what has and has not worked within insider threat mitigation strategies. Ultimately, published works from Park, You, and Lee (2018) led this dissertation study to further examine sentiment exposed in social media.

Research Question 2 asked: Was there relevance in behavioral theories, court transcripts from fraudulent cases, and social inputs that can solve the problem with the research? First, the existences of different behavioral theories were discovered within

each lab experiment, all aligned to fraud-related court cases, and shared one common behavioral theory; RAT (see Figure 43). In experiment one, the defendant situated himself into the trusted CEO position without mechanisms in place to prevent the embezzlement, and waited for the opportune time to execute the fraudulent act. In subsequent experiments, the defendants defrauded investors with various schemes and included fictitious businesses establishing collaborators such as managers, brokers, and processors to cover all aspects of the crime, all without protectors. In another instance, a contractor attempted to keep property belonging to a bank, and in all instances appeared to demonstrate an alignment to criminal acts requires convergence in space and time of likely offenders, suitable targets, and the absence of capable guardians against crime (Cohen & Felson, 1979).

Second, the lessor behavioral theory; Theory of Planned Behavior appeared in experiments two, three, and four to demonstrate the insufficiency of following any type of behavioral control, antecedents of attitudes, subjective norms, and perceived behavioral control that leads to predictors with intentions and actions (Ajzen, 1991). Furthermore, the Theory of Reasoned Action appeared in experiment four, and the Protection Motivation Theory appeared in experiment five (see Appendix X). In these instances, it appears through the knowledge of the preceding behavioral theories, there are likely ways organizations could mitigate the insider threat by implementing strategies to predict what people might do.

Third, just as behavioral theories drawn from fraudulent court cases showed a significant value, the collecting of social input data from Twitter tweets provided supplemental inputs. The AffectiveTweets package scored with Bravo et al.'s (2015)

NRC-Hash-Sent-negScore, Baccianella et al.'s (2010) SentiWordnet-negScore, Bravo's AFINN-negScore, and Bandhakavi et al.'s (2018) S140-negScoreAFIN (see Appendix K) to demonstrate relevancy within each experiment's output (see Table 10).

Table 10
AffectiveTweets Scores

Attribute	Description	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6
NRC-Hash-Sent-negScore	Sadness, anger, fear, and disgust	-1.244	-4.990	-4.990	-4.990	-4.990	-4.990
SentiWordnet-negScore	Top ranked negative synsets	-0.344	-0.291	-1.067	-0.208	-1.696	-1.067
AFINN-negScore	Slang, obscene words, web jargon	-3.000	-1.174	-4.000	-4.000	-4.000	-4.000
S140-negScore	Emotion aware negative tweets	-0.180	-0.180	-2.148	-4.990	-2.410	-2.148

In summary, there appears to be relevance in mitigating insider threats through the use of negative sentiment associated within the fraudulent context of social media. Furthermore, the discovery of repetitive behavioral theories might imply preventive measures to possible insider threats are probable.

Research Question 3 asked: What behavioral theories are most applicable to the research? Out of all behavioral theories, this dissertation study found the Routine Activity Theory common between the lab experiments. This research concurs with works by Cohen and Felson (1979) and the culmination of a suitable target, motivated offender, and the lack of an authority figure, allows criminal behaviors to develop. In every fraudulent case, Cohen and Felson's theory is in alignment with the events leading to the actions of the insider threat (see Appendix Y).

Research Question 4 asked: Can IT artifacts be created from the information obtained in behavioral theories, from court transcripts of fraudulent cases, and social inputs? The answer to this question is yes. First, after collecting court transcripts (see Table 11) from cases centered around fraudulent activities and documented in 138 pages of material, each case provided unique data that was paired against Hutto and Gilbert's (2014) rule-based model for sentiment analysis. Second, each of the lab experiments leveraged the Weka package *AffectiveTweets.* According to Bravo-Marquez et al. (2019), the package is used to analyze sentiment found in the 20,000 social media Twitter tweets and was instrumental in gathering *AffectiveTweets* scoring (see Table 10). Scores with the lowest negative values reflect sentiment intensity, and appear to be useful when identifying a potential threat within tweets. Third, applicable behavioral theories were annotated by observations throughout (see Appendix Y) with correlations connecting fraudulent behaviors to associated theories.

Table 11
Referenced Court Documents

Court	Assigned case number
The Superior Court of California. County of Santa Clara	SC-1903821
The United States District Court, Northern District of California, San Francisco Division	320-CR-00266
The United States District Court, Northern District of California, San Francisco Division	320-CR-00245
The United States District Court, Northern District of California, San Francisco Division	4-15-CV-01490
The United States District Court, Northern District of Illinois	3-16-CV-02600
The United States District Court, Southern District of California. The Southern District	3-16-CV-01547

Lastly, each IT artifact's creation began with Offermann et al.'s (2009) overarching design science research methodology to contribute to the uniform creation of each artifact (see Table 12) through the GPFLE process as outlined in the *Result's* section. Furthermore, the IT artifact's granular design leveraged works by Leoz and Petter (2018) to provide social aspects, behavioral aspects, and fraud-related aspects. Conversely, the technical component included information aspects, technology aspects, and an IT design, supportive of the IT artifact creation.

Table 12
Artifact Details

	Artifact 1	Artifact 2	Artifact 3	Artifact 4	Artifact 5	Artifact 6
Applicable behavioral theories	1	2	2	2	2	1
Pages of court transcripts	1	22	23	36	33	23
Negative lexicons from tweets	157	206	1846	5926	7515	3269

Research Question 5 asked: Will each IT artifact yield favorable outcomes through lab experiments and contribute to the goal of the study? All IT artifacts were based on Offermann et al.'s (2009) DSR model and implemented Leoz and Petter's (2018) artifact design with the majority of lab experiments producing what appears to be an expected outcome. The first lab experiment was limited in its data collections from court transcripts, and the artifact did not provide a significant outcome related to negative lexicons (see Appendix Q). However, in this instance, sufficient data did exist in order to correlate a behavioral theory (see Appendix Y) and reveal negative sentiment. For example, the AFINN score produced a -3.0 value and indicates the negative sentiment within the AFINN scale. Furthermore, the NRC attributes were derived from the word-level emotion association lexicon for about 14,200 word types (Mohammad & Turney, 2013) and produced a low HASH-SENT-negScore of -1.244.

Other experiments contributed to the study by identifying lexicons, behavioral theories (see Appendix Y), and sentiment scores likely associated with fraudulent tweets (see Table 10). These sane experiments extended data collections and provided 206, 1,846, 5,926, 7,515, and 3,269 classification instances for experiments two, three, four, five, and six. The dissertation study's promising outcomes arrived from the review of each set of court documents (see Table 11), established negative associations that correlated to various tweets, effectively scored overall sentiment using AffectiveTweets to examine social media tweet sentiment, and applied Hutto's (2014a) negative lexicons to demonstrate undesired conversations within tweets. Lastly, all outcomes appear favorable to improving the preemployment vetting process.

Implications

The study demonstrated that sentiment found in social media data could illuminate negativity associated with different flavors of fraud. The various IT artifacts' foundations were based on the unique cases on fraud and could be used by during the vetting process for employees, contractors, or business associates. Moreover, these research findings could contribute to a holistic solution to help mitigate insider threats and contribute to the body of knowledge.

Recommendations

First, as mentioned in the study's weaknesses, future research should explore support for additional languages and go beyond English (with permission of the author). Verhoeven, Daelemans, and Plank (2016) collected a sample of 65,000,000 tweets and found that Japanese, Spanish, Portuguese, Arabic, and French are used most frequently after English (see Figure 45). An insignificant portion of the Twitter data in the present study reflected tweets in other languages. For instance, out of twenty-thousand tweets, Spanish accounted for twenty tweets, Estonian accounted for one tweet, French accounted for two tweets, Italian accounted for three tweets, Dutch accounted for five tweets, Portuguese accounted for three tweets, and Tagalog accounted for twelve tweets.

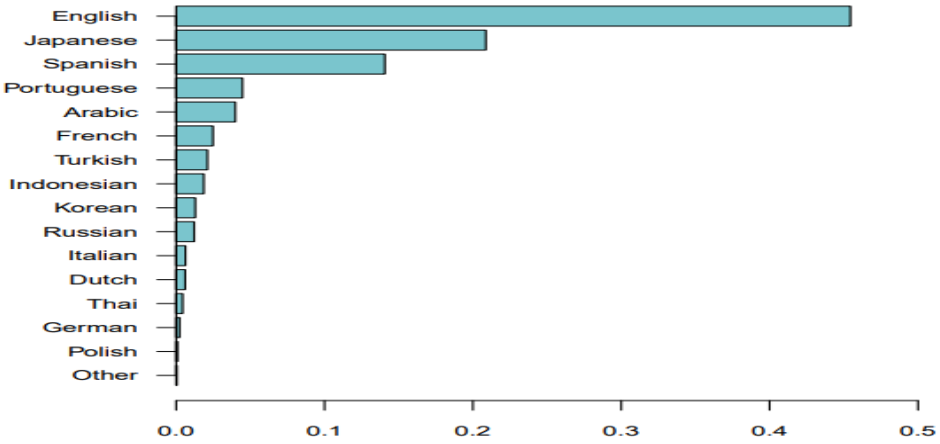


Figure 45. Distribution of languages (% of Tweets). Reprinted from "Twisty: A multilingual twitter stylometry corpus for gender and personality profiling," by B. Verhoeven, W. Daelemans, and B. Plank, 2016, Proceedings of the Tenth International Conference on Language Resources and Evaluation, 1632–1637. Adapted with permission.

Second, further expansion beyond fraud is needed to target specific forms of white-collar crimes, as previously mentioned. Expanding to other forms of crime could yield a wealth of data and benefit organizations' mitigation efforts.

Furthermore, other social media tools could be leveraged. This dissertation study did not realize many of the initial up-front and technical challenges; thus, dealt with problematic issues in on-premises processing and analyzing data for various technological reasons. A later discovery led to more tools with many of companies offering trial periods, discounted pricing, and some required purchasing. However, due to cloud privacy concerns, caution should be used when using products that are eager to place data in cloud hosting environments. The following (see Table 13) are a list of additional tools that were discovered during the research process:

Table 13
Available Tools

Tool	Description
Comprehend	Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to discover insights from text (Amazon.com, 2021).
TalkWalker	AI powered analysis provides real time insights into what's happening on all social channels and online media, across 187 languages (Talkwalker.com, 2021).
Tableau Desktop	Visual analytics displays interactive dashboards help you uncover hidden insights on the fly (Tableau.com, 2021).
DataRobot	DataRobot is the leading end-to-end enterprise AI platform that automates and accelerates every step of your path from data to value (DataRobot.com, 2021).
RapidMiner	According to RapidMiner.com (2021), the product provides a comprehensive data science platform to utilize automation and visual workflow design.
MLbase	According to MLbase.org (2021), MLbase is distributed machine learning consisting of three components. *MLlib*, *MLI*, and *ML Optimizer*, to address issues with implementing and consuming machine learning tasks.
BigML	BigML is a consumable, programmable, and scalable Machine Learning platform that makes it easy to solve and automate Classification, Regression, Time Series Forecasting, Cluster Analysis, Anomaly Detection, Association Discovery, and Topic Modeling tasks (BigML.com, 2021).
Datawrapper	Enrich your stories with charts, maps, and tables (Datawrapper.de, 2021).
Visualr	Visualr is a Data Visualization and Analytics Platform that will help your organization to convert raw data into insights in the form of interactive Dashboards and Analytical Reports, from different data sources (Visualr.io, 2021), and is capable of handling data in the terabytes.
Paxata	Paxata provides a self-service data preparation solution for business and technical teams to visually clean, integrate, and govern data at scale (Paxata.com, 2021).
Trifacta	Trifacta provides visual and intelligent guidance to accelerate data preparation so you get to insights faster (Trifacta.com, 2021).

Summary

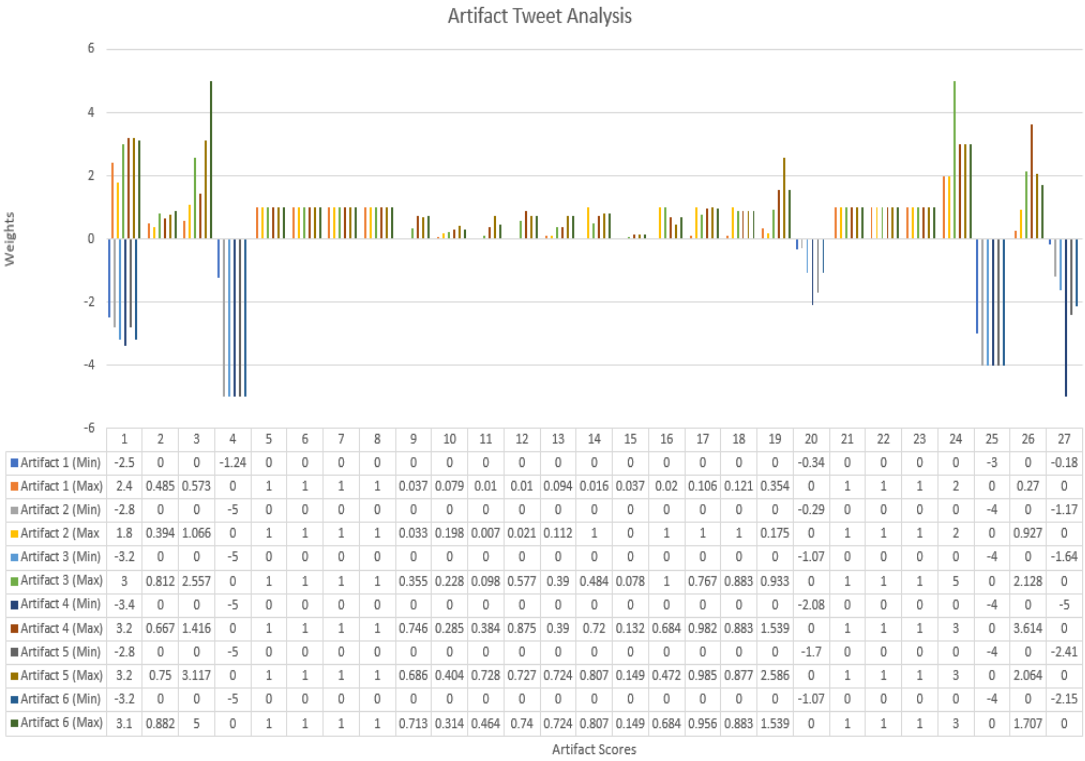
The opening chapter introduced the background of insider threats through an extensive examination of past, current, and future directions of insider threat activities. Previous postures demonstrated a reactive stance, which often require additional personnel and technology support (Wallace & Loffi, 2014). This dissertation study closely explored the social media sentiments presented by Gritzalis et al. (2014), who postulated that online content provides characteristics of individuals who demonstrate traits of a potential insider threat.

This dissertation study applied 11 steps from a DSR methodology that led to the creation of IT artifacts. Multiple sources of data collections, analysis, instrument development, and validation provided the sentiment scores and negative sentiment classifications that contributed to identifying insider threats from an earlier stage within social media data. Similarly, research by Park et al. (2018) included social media data and behavioral theories, which provided the foundation for this study. Such information contributed to the relevancy of moving towards proactive measures.

Several theories were more applicable than others. For instance, the general deterrence theory and the social bond theory did not appear to be associated with any of the court input streams. However, the protection motivation theory, TRA, the RPB, and the RAT appeared to be associated with the same set of cases, with the most prominent being the RAT and the TPB. The latter two theories may have been most prominent simply because of similarities within the court cases.

The summary of results below illustrates that all artifact Tweet analyses were paired with negative sentiments extracted from the various forms of fraudulent court

cases. Although this dissertation study did anticipate retrieving positive sentiments, it did not show bias by restricting such data. Instead, focus on negative sentiments led to the prediction of insider threats. Within the artifact Tweet analysis, all Tweet filtering measures for each artifact are listed in columns 1–27, with the lowest negative score representing the point of interest as shown in columns 1, 4, 20, and 25 (see Figure 46).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Artifact 1 (Min)	-2.5	0	0	-1.24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.34	0	0	0	0	-3	0	-0.18
Artifact 1 (Max)	2.4	0.485	0.573	0	1	1	1	1	0.037	0.079	0.01	0.01	0.094	0.016	0.037	0.02	0.106	0.121	0.354	0	1	1	1	2	0	0.27	0
Artifact 2 (Min)	-2.8	0	0	-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.29	0	0	0	0	-4	0	-1.17
Artifact 2 (Max	1.8	0.394	1.066	0	1	1	1	1	0.033	0.198	0.007	0.021	0.112	1	0	1	1	1	0.175	0	1	1	1	2	0	0.927	0
Artifact 3 (Min)	-3.2	0	0	-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.07	0	0	0	0	-4	0	-1.64
Artifact 3 (Max)	3	0.812	2.557	0	1	1	1	1	0.355	0.228	0.098	0.577	0.39	0.484	0.078	1	0.767	0.883	0.933	0	1	1	1	5	0	2.128	0
Artifact 4 (Min)	-3.4	0	0	-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-2.08	0	0	0	0	-4	0	-5
Artifact 4 (Max)	3.2	0.667	1.416	0	1	1	1	1	0.746	0.285	0.384	0.875	0.39	0.72	0.132	0.684	0.982	0.883	1.539	0	1	1	1	3	0	3.614	0
Artifact 5 (Min)	-2.8	0	0	-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.7	0	0	0	0	-4	0	-2.41
Artifact 5 (Max)	3.2	0.75	3.117	0	1	1	1	1	0.686	0.404	0.728	0.727	0.724	0.807	0.149	0.472	0.985	0.877	2.586	0	1	1	1	3	0	2.064	0
Artifact 6 (Min)	-3.2	0	0	-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1.07	0	0	0	0	-4	0	-2.15
Artifact 6 (Max)	3.1	0.882	5	0	1	1	1	1	0.713	0.314	0.464	0.74	0.724	0.807	0.149	0.684	0.956	0.883	1.539	0	1	1	1	3	0	1.707	0

Figure 46. Artifact Tweet analysis.

Additionally, some instances represented extreme levels of negative sentiment which were classified under sentiment weights, such as NRC-Hash-Sent-negScores, SentiWordnet-negScores, and AFINN-negScores. Recognizing these types of sentiment scores as applied to social media analysis could expose fraud-related insider threats during preemployment vetting. The negative sentiments could provide HR with another tool to mitigate potential threats within an organization and its business partners.

Appendices

Appendix A:

Fix Sentiment Weights

```
Module FixSentimentWeights
'*********************************************************************************************
'* Program: FixSentimentWeights
'*       Date: 06-09-2020
'*         By: Robert W. Jones
'*   Purpose: Create a new source file from the original VADER lexicon file to correct
'*              issues with either missing weights and deleted a record at position 825
'*              that appeared to cause issues when reading. The output file has every
'*              piece of information verbatim to the master input. To correct an issue
'*              that caused WEKA to error when reading, the output from this program did
'*              require opening the new output file in NotePad, then as an ANSI file (as
'*              the original used UTF-8 encoding.
'*********************************************************************************************

Sub Main()
Dim sSource As String = "" 'Source filename
Dim sDestination As String = "" 'Destination filename
Dim sRecordOut As String = "" 'String record out
Dim sRecordIn As String = "" 'String record in
Dim nStartPos As Byte = 0 'Numeric start position
Dim sOut As String = "" 'String out
Dim nFilePosition As Integer = 0 'Numeric file positioning

sSource = "C:\DISS901-3\vaderSentiment-master\vaderSentiment\vader_lexicon2.arff"
sDestination = "C:\DISS901-3\vaderSentiment-master\vaderSentiment\vader_lexicon_gold.arff"

Dim file As System.IO.StreamWriter 'Outfile
file = My.Computer.FileSystem.OpenTextFileWriter(sDestination, False)
Try

If System.IO.File.Exists(sSource) = True Then
Dim objReader As New System.IO.StreamReader(sSource)
Do While objReader.Peek() <> -1
sRecordIn = objReader.ReadLine()
nFilePosition += 1
If nFilePosition <= 6 Then 'No changes until after record 6
file.WriteLine(sRecordIn)
Else
'After record 6, then process and make changes in destination file
nStartPos = InStr(sRecordIn, ",", CompareMethod.Text) + 1
sOut = Mid(sRecordIn, nStartPos, 4)
file.WriteLine(RTrim(sRecordIn) + ",{" + sOut + "}")
End If
Loop 'Read all VADER data that is available
Else
```

Appendix A continued:

Fix Sentiment Weights

```
MsgBox("Error opening " & sSource)
Exit Sub 'Unable to open the source VADER data file
End If

'Close file
file.Close()

Catch ex As Exception
MsgBox("FixSentimentWeights has encountered an error and unable to continue.")
End Try

End Sub

End Module
```

Appendix B

Background on VADER lexicons

VADER was empirically validated using multiple and independent judges when establishing the “gold-standard” with sentiment that leverages blog-like contexts. Lexicons implements both polarity and intensity of sentiments tuned to social media. According to C.J. Hutto (2014),

> Sentiment ratings from 10 independent human raters (all pre-screened, trained, and quality checked for optimal inter-rater reliability). Over 9,000 token features were rated on a scale from "[–4] Extremely Negative" to "[4] Extremely Positive", with allowance for "[0] Neutral (or Neither, N/A)". We kept every lexical feature that had a non-zero mean rating, and whose standard deviation was less than 2.5 as determined by the aggregate of those ten independent raters. This left us with just over 7,500 lexical features with validated valence scores that indicated both the sentiment polarity (positive/negative), and the sentiment intensity on a scale from –4 to +4. For example, the word "okay" has a positive valence of 0.9, "good" is 1.9, and "great" is 3.1, whereas "horrible" is –2.5, the frowning emoticon :(is –2.2, and "sucks" and it's slang derivative "sux" are both –1.5. Manually creating a comprehensive sentiment lexicon is a labor intensive and sometimes error prone process, so it is no wonder that many opinion mining researchers and practitioners rely so heavily on existing lexicons as primary resources. We are pleased to offer ours as a new resource. We began by constructing a list inspired by examining existing well-established sentiment word-banks (LIWC, ANEW, and GI). We empirically confirmed the general applicability of each feature candidate to sentiment expressions using a wisdom-of-the-crowd (WotC) approach (Surowiecki, 2004) to acquire a valid point estimate for the sentiment valence (polarity & intensity) of each context-free candidate feature. (p. 1)

Appendix C

Import Social Media into WEKA

Figure C1

WEKA Explorer

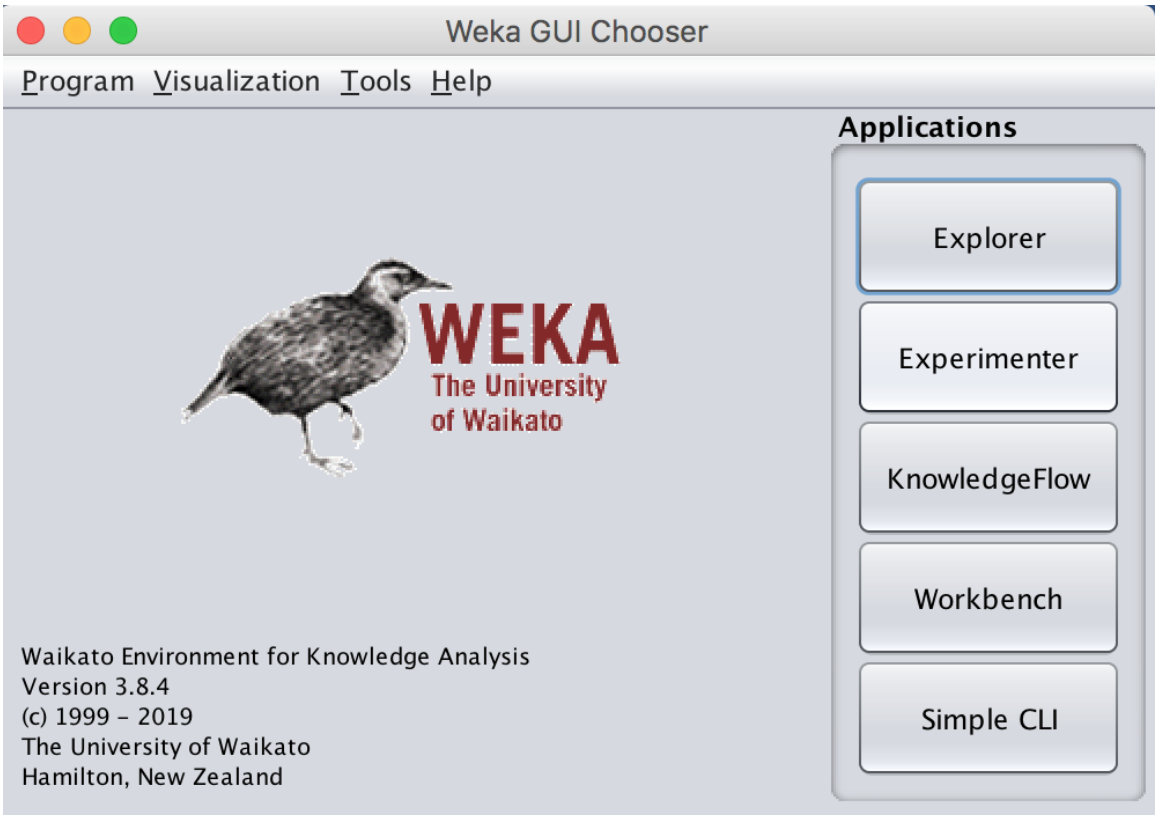


Figure C2

Tweet ARFF File

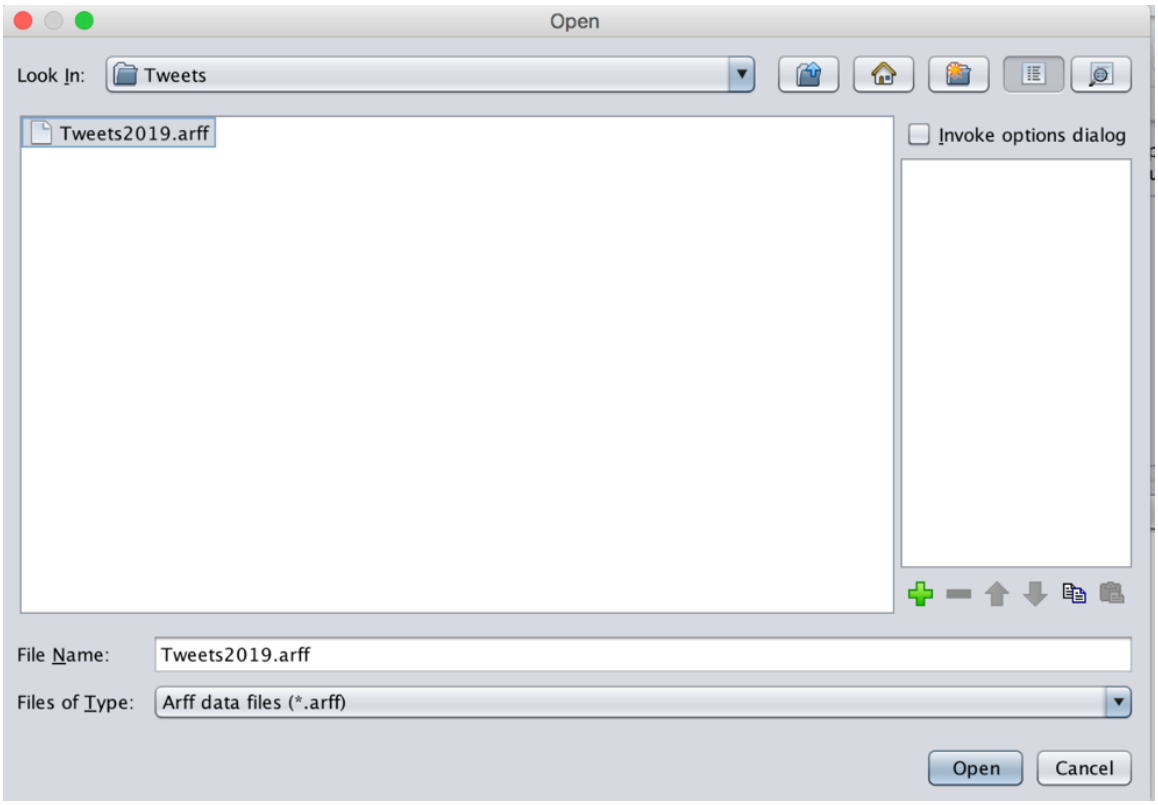


Figure C3

Tweet Filtering

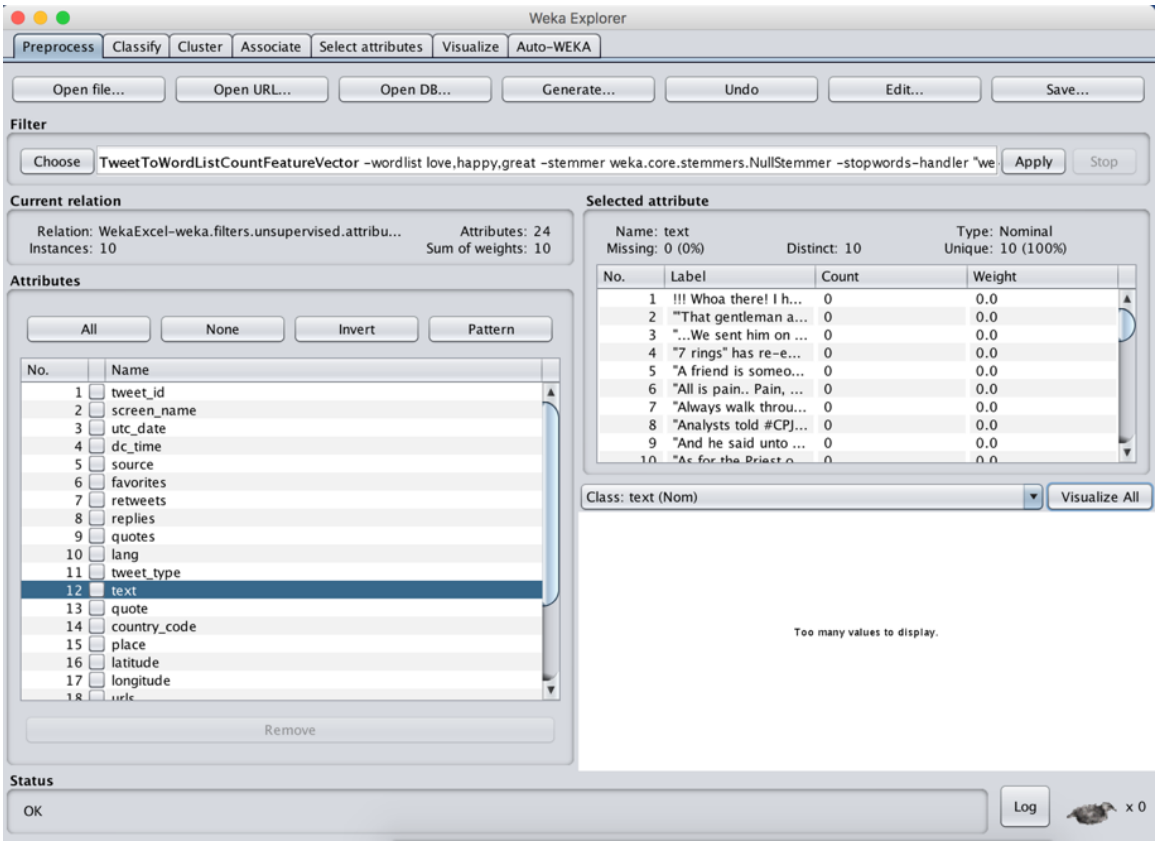
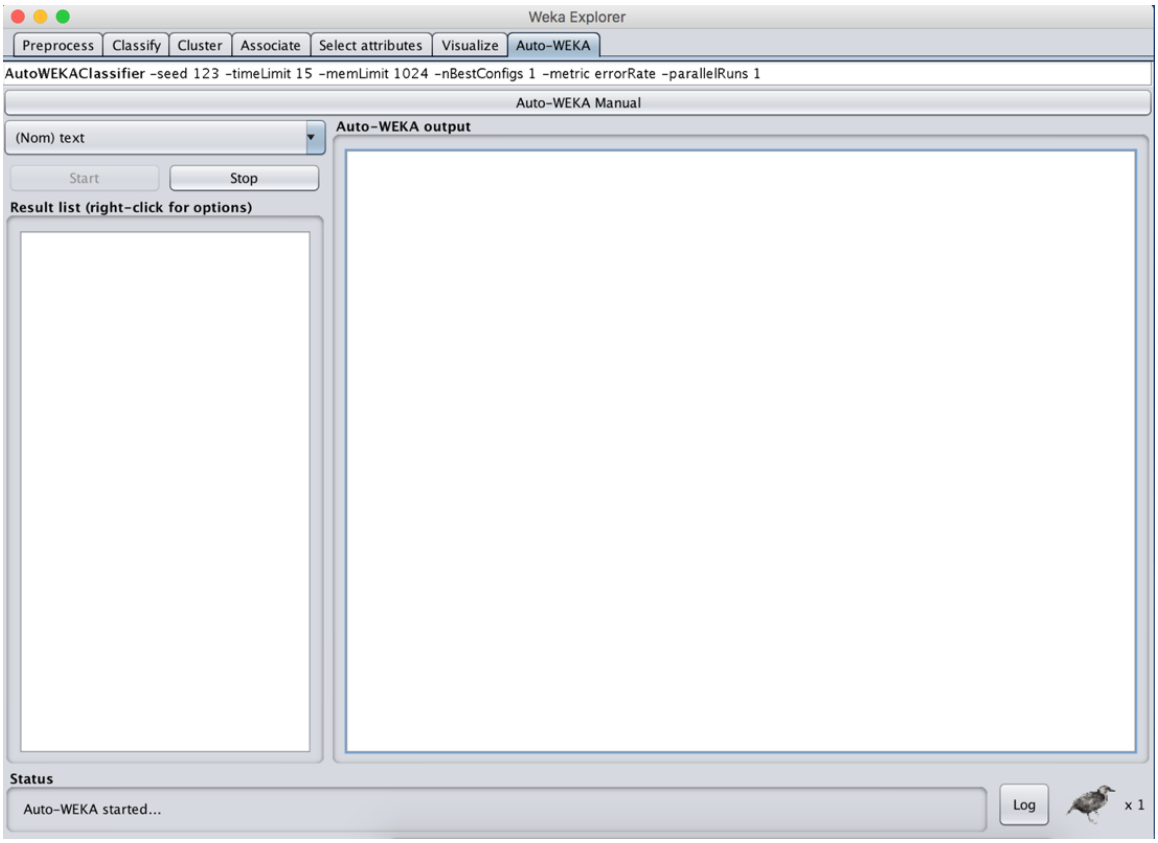


Figure C4

Auto-WEKA Sentiment Analysis



Appendix D

Illustrations of Sentiment Analysis for Input and Output

Figure D1

Illustrations of Sentiment Analysis for Input and Output (caption 1)

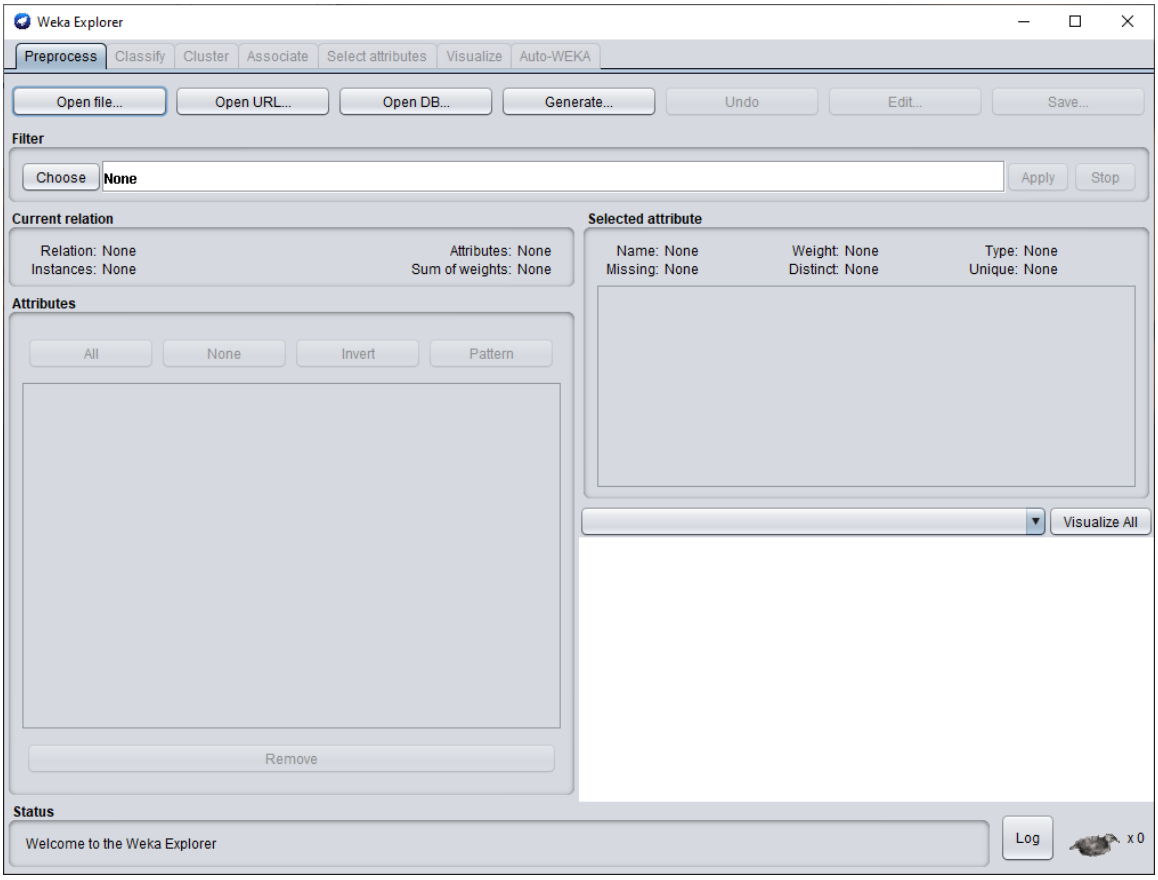


Figure D2

Illustrations of Sentiment Analysis for Input and Output (caption 2)

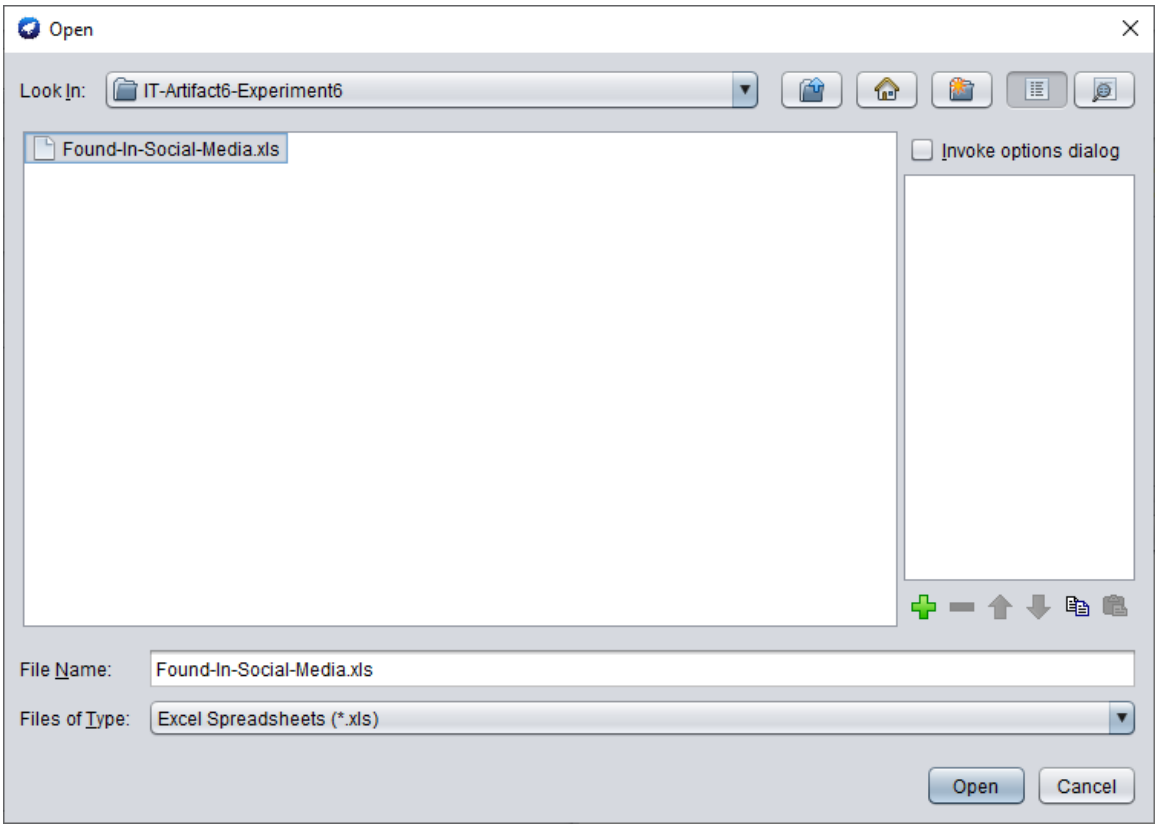


Figure D3

Illustrations of Sentiment Analysis for Input and Output (caption 3)

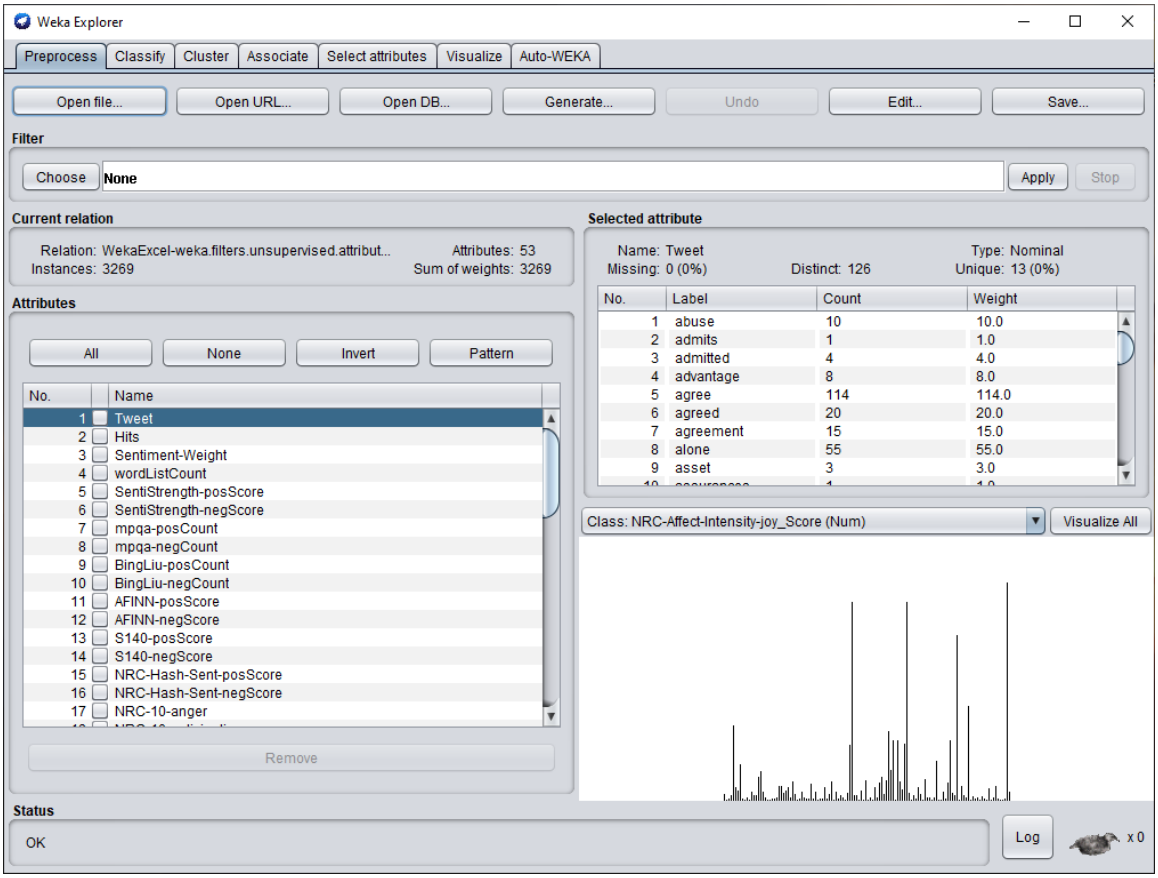


Figure D4

Illustrations of Sentiment Analysis for Input and Output (caption 4)

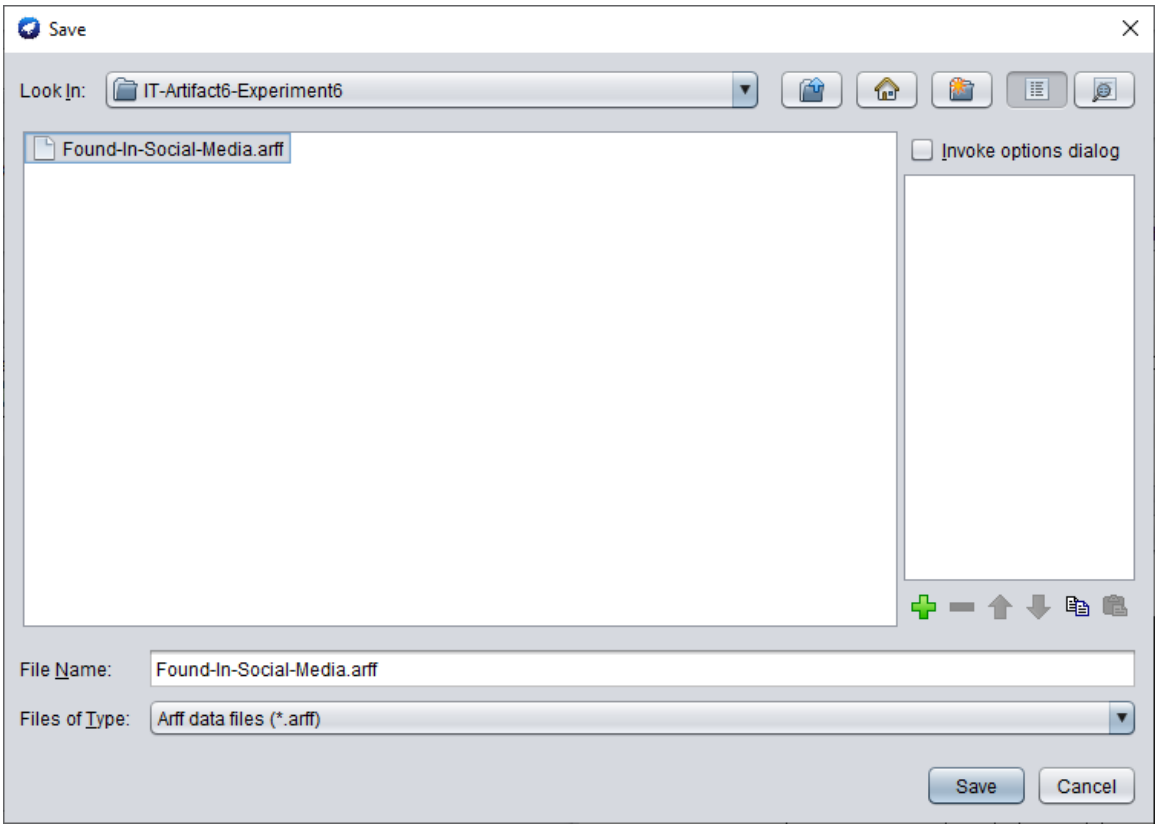


Figure D5

Illustrations of Sentiment Analysis for Input and Output (caption 5)

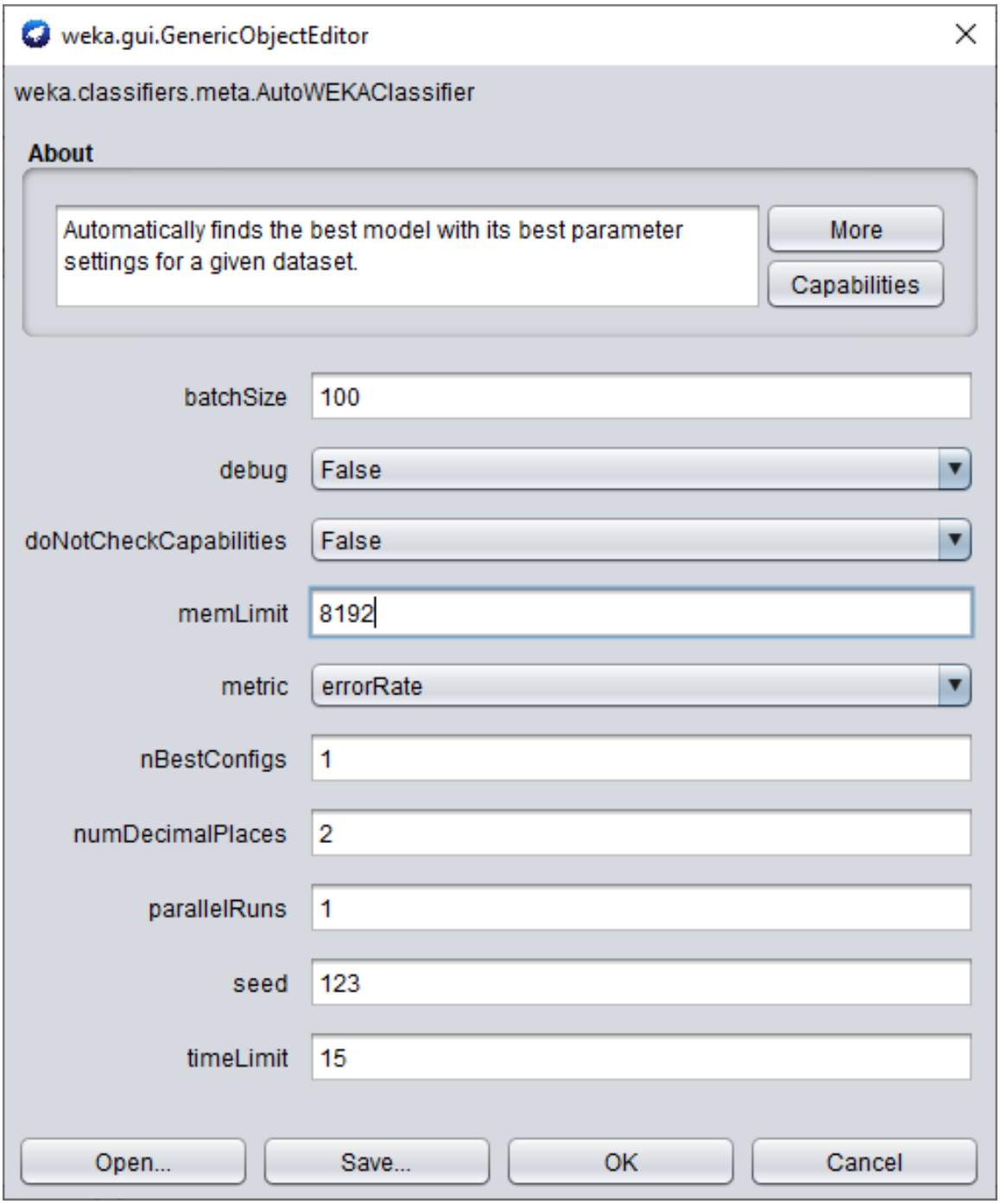


Figure D6

Illustrations of Sentiment Analysis for Input and Output (caption 6)

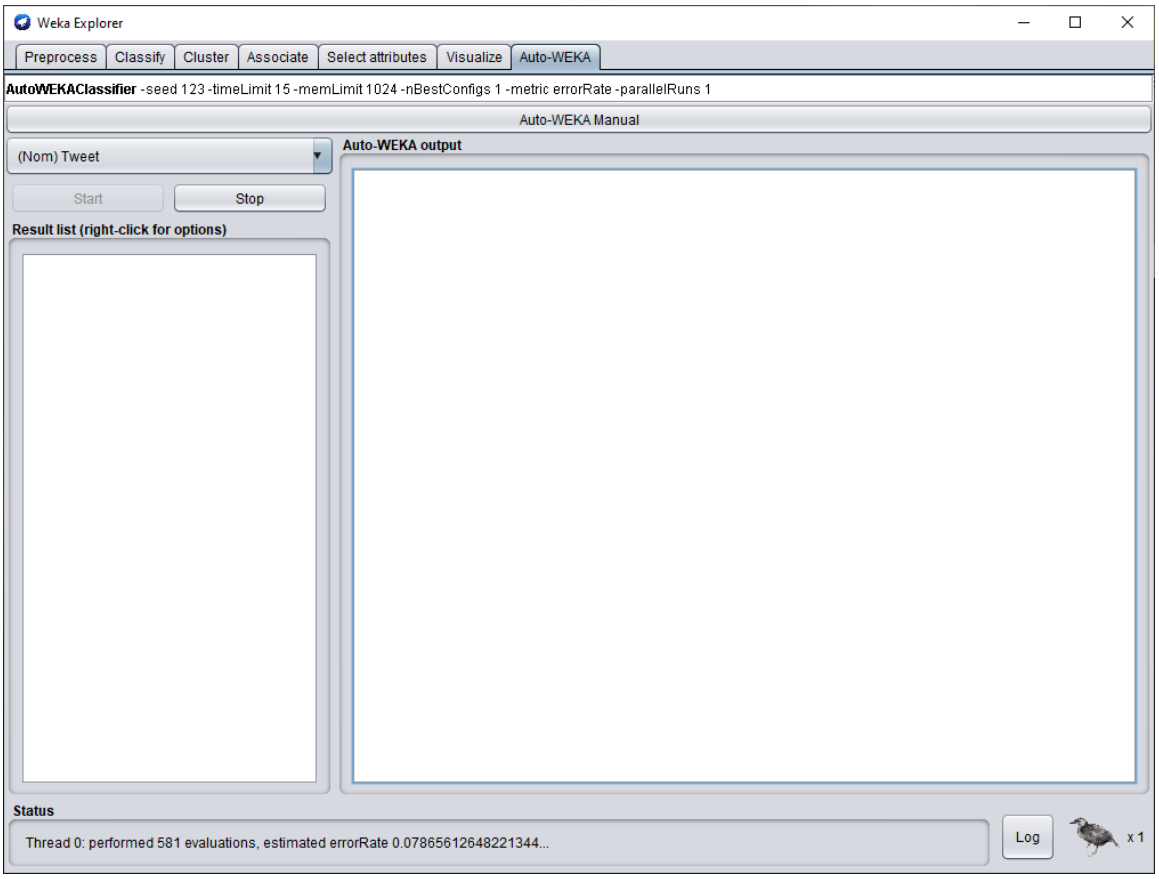


Figure D7

Illustrations of Sentiment Analysis for Input and Output (caption 7)

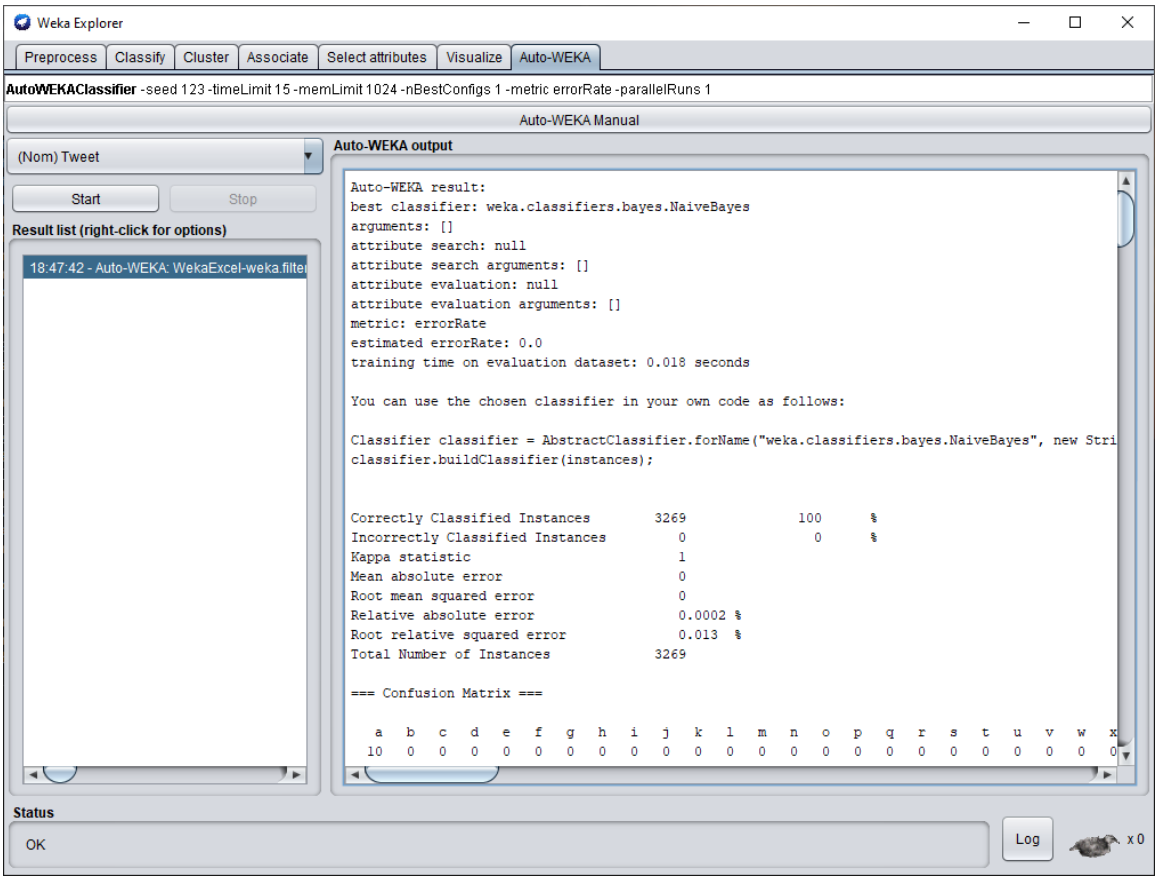
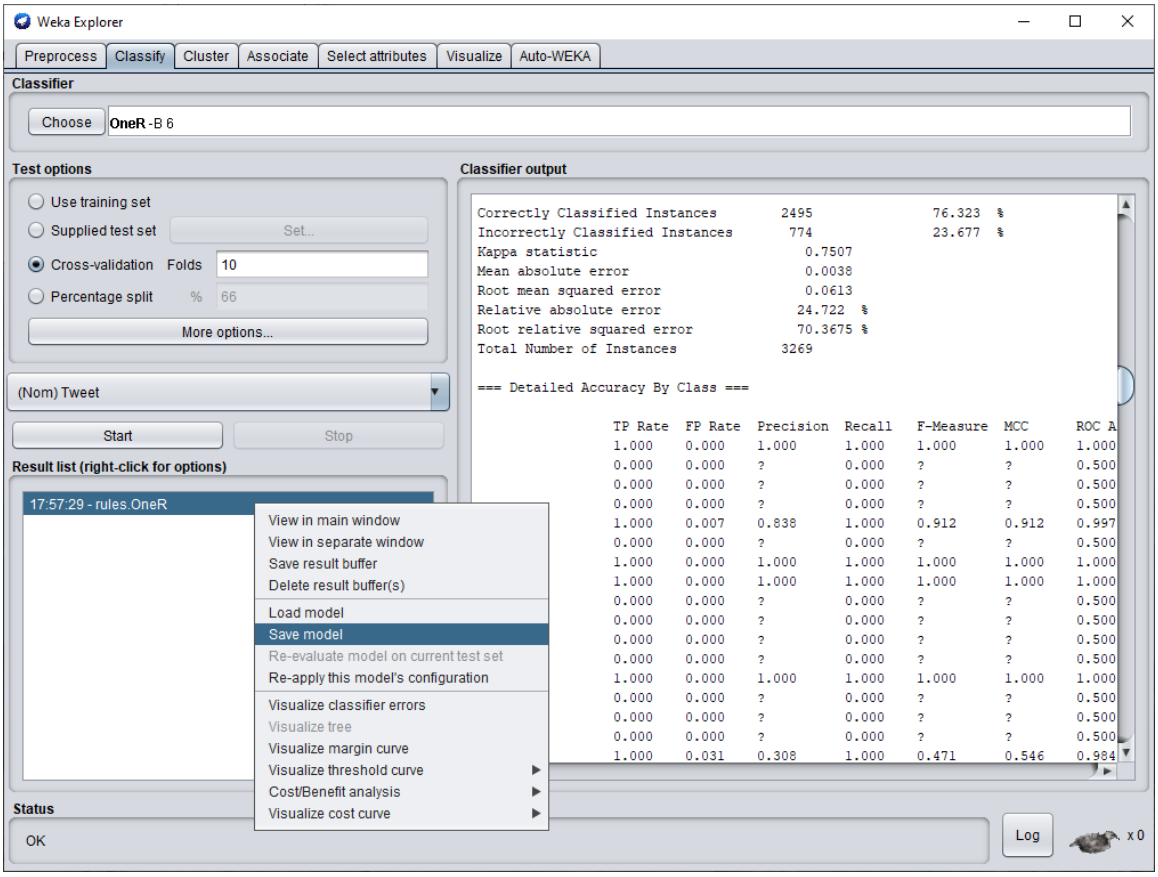


Figure D8

Illustrations of Sentiment Analysis for Input and Output (caption 8)



Appendix E

Assign Weights To Court Lexicons

```
Module Module1
Sub Main()
'*************************************************************************************
'* Program: AssignWeightsToCourtLexicons
'*      Date: 06-09-2020
'*        By: Robert W. Jones and AutomateExcel.com use of modified extact_number()
'*  Purpose: After PDFs are converted to Excel, then sorted and saved to TEXT,
'*           this program reads the TEXT file, and creates a new output file
'*           with the content from the input, and paired with associate sentiment
'*           weight from C.J. Hutto (2014) sentiment weights. The output should
'*           be read into WEKA, along with Twitter data and have sentiment analysis
'*           performed.
'*************************************************************************************
Dim sLexicon(7065) As String 'Array to store sentiment lexicons
Dim sWeight(7065) As String 'Array to store sentiment weights
Dim nPointer As Integer = 0 'Integer pointer
Dim sSource As String = "" 'Source filename
Dim sDestination As String = "" 'Destination output filename
Dim sSourceSentiment As String = "" 'Source sentiment filename
Dim sRecordOut As String = "" 'Record out
Dim sRecordIn As String = "" 'Record in
Dim stemp1 As String = "" 'Temp string
Dim stemp2 As String = "" 'Temp string
Dim bFound As Boolean = False 'Booleam for finding match
Dim nCtr As Integer = 0 'Integer for processing array
Dim nFreeFile1 As Integer 'File handle
Dim file As System.IO.StreamWriter 'Outfile handle for writing

Try

'Source file was created from using iSkySoft to convert the court PDF file(s) into a text file that was
cleaned up from duplicates using Excel, and sorted in ascending order.
sSource = "C:\DISS901-3\Artifact Lab Results\CA-C1903821.TXT"

'Destination file uses the sSource and rewrites it, so the sentiments weighted value is included. If no
sentiment weight is found, a zero is assigned.
sDestination = "C:\DISS901-3\Artifact Lab Results\CA-C1903821-ReadyForWEKA.TXT"
'Source file for reading sentiment values
sSourceSentiment = "C:\DISS901-3\vaderSentiment-master\vaderSentiment\vader_lexicon_txt.txt"

'Open the VADER sentiment as sSourceSentiment (Read Only)
'Read into two dim array (for faster processing)
'Close the VADER sentiment sSourceSentiment file

nFreeFile1 = FreeFile()
```

Appendix E continued:

Assign Weights To Court Lexicons

```
Dim objReader As New System.IO.StreamReader(sSourceSentiment)
Do While objReader.Peek() <> -1
sRecordIn = objReader.ReadLine()
'Pull values into arrays
stemp1 = Mid$(sRecordIn, 1, InStr(1, sRecordIn, vbTab, CompareMethod.Text) - 1) 'This is working and only pulling the sentiment.
stemp2 = Extract_Number(sRecordIn) 'This is working and only pulling weights.
sLexicon(nPointer) = stemp1
sWeight(nPointer) = stemp2
nPointer += 1
Loop

'Close sentiment file when done reading into array
FileClose(nFreeFile1)

'Open the output file for writing.
file = My.Computer.FileSystem.OpenTextFileWriter(sDestination, False)

'Open the court data file as sSource (Read only)
Dim objReadersSource As New System.IO.StreamReader(sSource)
Do While objReadersSource.Peek() <> -1
bFound = False
sRecordIn = objReadersSource.ReadLine()
'Search array to find weight. If found, write to output file with new weight. If not found, write to new file with a zero weight.
For nCtr = 0 To 7065 Step 1
If RTrim(sRecordIn) = RTrim(sLexicon(nCtr)) Then
bFound = True
Exit For 'No need to keep looking once a match is found
End If
Next
'Write whichever is needed
If bFound = True Then
file.WriteLine(RTrim(sRecordIn) & " " & sWeight(nCtr)) 'Write matched weight
Else
file.WriteLine(RTrim(sRecordIn) & " " & "0.0") 'Write zero as no match in sentiment master file (not good/nor bad)
End If
Loop

file.Close()
Catch ex As Exception
MsgBox("AssignWeightsToCourtLexicons has encountered an error and unable to continue.")

 End Try
 End Sub
```

Appendix E continued:

Assign Weights To Court Lexicons

```
Function Extract_Number(Phrase As String) As String
Dim Length_of_String As Integer
Dim Current_Pos As Integer
Dim Temp As String
Length_of_String = Len(Phrase)
Temp = ""
For Current_Pos = 1 To Length_of_String
If (Mid(Phrase, Current_Pos, 1) = "-") Then
Temp = Temp & Mid(Phrase, Current_Pos, 1)
End If
If (Mid(Phrase, Current_Pos, 1) = ".") Then
Temp = Temp & Mid(Phrase, Current_Pos, 1)
End If
If (IsNumeric(Mid(Phrase, Current_Pos, 1))) = True Then
Temp = Temp & Mid(Phrase, Current_Pos, 1)
End If

Next Current_Pos
If Len(Temp) = 0 Then
Extract_Number = 0
Else
Extract_Number = Temp
End If
End Function
End Module
```

Appendix F

Convert Twitter Tweets into Fixed Words

```
‘'*********************************************************************************
 '* Program: ConvertTwitterToWords
 '*       Date: 06-16-2020
 '*         By: Robert W. Jones
 '*              Purpose: Read the 20,000 tweets from the master CSV file and convert into plain text.
 '*
 '*********************************************************************************
 'Sub Main()
  Dim sSource As String = "" 'Source filename
  Dim sDestination As String = "" 'Destination filename
  Dim sRecordOut As String = "" 'String record out
  Dim sRecordIn As String = "" 'String record in
  Dim sOut As String = "" 'String out
  Dim nCtr As Integer = 0 'For loop counter

  sSource = "C:\DISS901-3\Tweets\TwitterMasterGold-CSV.csv"
  sDestination = "C:\DISS901-3\Tweets\Twitter.txt"
  Dim file As System.IO.StreamWriter 'Outfile
  file = My.Computer.FileSystem.OpenTextFileWriter(sDestination, False)

  Try
  If System.IO.File.Exists(sSource) = True Then
  Dim objReader As New System.IO.StreamReader(sSource)
  Do While objReader.Peek() <> -1
   sRecordIn = objReader.ReadLine()
   sOut = "" 'Reinitialize
   For nCtr = 1 To Len(RTrim(sRecordIn)) Step 1
   If Mid(sRecordIn, nCtr, 1) <> Space(1) Then
   If Mid$(sRecordIn, nCtr, 1) = "?" Then
   sOut = sOut & "'" 'Convert it as it causes issues
   ElseIf Mid$(sRecordIn, nCtr, 1) = "." Then
   sOut = sOut & " " 'Convert it as it causes issues
   ElseIf Mid$(sRecordIn, nCtr, 1) = "," Then
   sOut = sOut & " " 'Convert it as it causes issues
   ElseIf Mid$(sRecordIn, nCtr, 1) = Chr(34) Then
   sOut = sOut & " " 'Convert it as it causes issues
   Else
   sOut = sOut & Mid(sRecordIn, nCtr, 1) 'Build the word
   End If    Next
    Loop 'Read all input records from tweet master CSV file
    Else
    MsgBox("Error opening " & sSource)
    Exit Sub 'Unable to open the source Twitter tweets data file
    End If
    'Close file
    file.Close()
    Catch ex As Exception
     MsgBox("ConvertTwitterToWords has encountered an error and unable to continue.")
```

Appendix F continued:

Assign Weights To Court Lexicons

```
    End Try
End Sub
```

Appendix G

Sequencer 1

Sequencer1 reads each line from court documents(s), looks up Hutto matches and records into a HITS file; used in later analysis by Sequencer2.

```
'*********************************************************************************************
'* Program: Sequence1
'*       Date: 06-16-2020
'*         By: Robert W. Jones
'*  Purpose: Read each line from court document(s) - referred to as CCxxxxxxx.txt, and
'*             : look-up in Hutto.txt file. If match is found, write output file as Hits.txt
'*             : with matching sentiment and weight (i.e., trouble, -1.5). If match is not
'*             : found, write to same file with a neutral zero weight (i.e., word, 0.0).
'*********************************************************************************************
Dim sSourceC As String = "" 'Source filename for court document
Dim sSourceH As String = "" 'Source filename for Hutto file
Dim sDestination As String = "" 'Destination Hits file
Dim sRecordOut As String = "" 'String record out
Dim sRecordIn As String = "" 'String record in
Dim nStartPos As Byte = 0 'Numeric start position
Dim sOut As String = "" 'String out
Dim nFilePosition As Integer = 0 'Numeric file positioning
Dim nCtr As Integer = 0 'Used in array for loop
Dim sHuttoLexicon(7064) As String 'Hold Hutto lexicons
Dim bFound As Boolean = False 'Found a match
Dim sHuttoTemp As String 'Temp var to hold temp lexicon strings
Dim sCourtTemp As String 'Temp var to hold temp court strings
Dim nCourtRecs As Long = 0 'Represent all court records

sSourceC = "C:\DISS-Work\IT-Artifact1-Experiment1\CA-C1903821.txt"
sSourceH = "C:\DISS-Work\IT-Artifact1-Experiment1\Hutto.txt"
sDestination = "C:\DISS-Work\IT-Artifact1-Experiment1\Hits.txt"

'Place sentiments into an array for faster processing.
Dim objReaderH As New System.IO.StreamReader(sSourceH)
Do While objReaderH.Peek() <> -1 'Read entire court documents
nCtr = nCtr + 1
sRecordIn = objReaderH.ReadLine() 'Read line
sHuttoLexicon(nCtr) = sRecordIn 'Assign to array
Loop 'Read all Hutto lexicons data
objReaderH.Close()

Dim file As System.IO.StreamWriter 'Outfile
file = My.Computer.FileSystem.OpenTextFileWriter(sDestination, False)

Try

If System.IO.File.Exists(sSourceC) = True Then
Dim objReader As New System.IO.StreamReader(sSourceC)
Do While objReader.Peek() <> -1 'Read entire court document
```

Appendix G continued:

Sequencer1 reads each line from court documents(s), looks up Hutto matches and records into a HITS file; used in later analysis by Sequencer2.

```
bFound = False
sRecordIn = objReader.ReadLine() & Space(1) 'Helps reduce false hits (i.e., court is not scored to courteous)
'See if match is found in Hutto array
For nCtr = 1 To 7064 Step 1

'Attempt to make each words ready for comparisons
sCourtTemp = LCase(RTrim(sRecordIn))
sHuttoTemp = LCase(RTrim(Replace(sHuttoLexicon(nCtr), vbTab, " ")))
sHuttoTemp = RTrim(Mid(sHuttoTemp, 1, InStr(1, sHuttoTemp, " ")))

'If InStr(1, LCase(sHuttoLexicon(nCtr)), LCase(RTrim(sRecordIn))) > 0 Then 'Match is found and not empty!
If (sCourtTemp = sHuttoTemp) Then
file.WriteLine(sHuttoLexicon(nCtr))
nCourtRecs += 1
bFound = True
'Exit For 'As soon as a match is found, move on to save time.
End If
Next
'If here and no match was found, write record in Hits with a zero weight.
If bFound = False Then
file.WriteLine(sRecordIn & Space(5) & "0.0")
End If
Loop 'Read all court data
Else
MsgBox("Error opening " & sSourceC)
Exit Sub 'Unable to open the source court data file
End If

'Close file
file.Close()
MsgBox("Data saved to " & sDestination)

Catch ex As Exception
MsgBox("Sequence1 has encountered an error and unable to continue.", ex.Message)
End Try
```

Appendix H

Sequencer 2

Sequencer2 reads each line from the HITS file and look-up possible hits in the Twitter master file. Recording the tweet hits, and weights. If matches are not found, then the lexicon is recorded with a zero weight.

```
'*****************************************************************************************
'* Program: Sequence2
'*       Date: 06-17-2020
'*         By: Robert W. Jones
'*  Purpose: Read each line from HITS and look-up possible hits in the Twitter file.
'*             : If matching tweets exists, record the lexicon and number of times
'*             : found (i.e., trouble, -1.5). If match is not found, write to same file
'*             : using a zero weight as 0.0.
‘*****************************************************************************************
Dim sSourceHits As String = "" 'Source filename for HITS document
Dim sSourceTwitter As String = "" 'Source filename for Twitter file
Dim sDestination As String = "" 'Destination Found-In-Social-Media file
Dim sRecordIn As String = "" 'String record in
Dim nCtr As Long = 0 'Used in array for loop
Dim sTweetedWords(382830) As String 'Hold Twitter tweets
Dim bFound As Boolean = False 'Found a match
Dim nTweetRecs As Long = 0 'Count matching tweets
Dim sTwitterTemp As String 'Temp var to hold temp twitter strings
Dim sHITsTemp As String 'Temp var to hold temp HITS strings
Dim nCourtRecs As Long = 0 'Represent all court records

sSourceHits = "C:\DISS-Work\IT-Artifact1-Experiment1\Hits.txt"
sSourceTwitter = "C:\DISS-Work\IT-Artifact1-Experiment1\Twitter.txt"
sDestination = "C:\DISS-Work\IT-Artifact1-Experiment1\Found-In-Social-Media.txt"

'Place large twitter words an array for faster processing.
Dim objReaderH As New System.IO.StreamReader(sSourceTwitter)
Do While objReaderH.Peek() <> -1 'Read entire twitter tweets
nCtr = nCtr + 1
sRecordIn = objReaderH.ReadLine() 'Read line
sTweetedWords(nCtr) = sRecordIn 'Assign to array
Loop 'Read all Hutto lexicons data
objReaderH.Close()

Dim file As System.IO.StreamWriter 'Outfile
file = My.Computer.FileSystem.OpenTextFileWriter(sDestination, False)
'Write a header for the csv and WEKA
file.WriteLine("Tweet" & vbTab & "Sentiment-Weight" & vbTab & "Hits" & vbLf)

Try

If System.IO.File.Exists(sSourceHits) = True Then
Dim objReader As New System.IO.StreamReader(sSourceHits)
```

Appendix H continued:

Sequencer2 reads each line from the HITS file and look-up possible hits in the Twitter master file. Recording the tweet hits, and weights. If matches are not found, then the lexicon is recorded with a zero weight.

```
Do While objReader.Peek() <> -1 'Read entire HITS document
bFound = False
nTweetRecs = 0
sRecordIn = objReader.ReadLine()

'See if match(es) are found in TWITTER file
For nCtr = 1 To 382830 Step 1
'Attempt to make each words ready for comparisons
sHITsTemp = LCase(RTrim(sRecordIn))
sHITsTemp = RTrim(Mid(sHITsTemp, 1, InStr(1, sHITsTemp, " ")))

sTwitterTemp = LCase(RTrim(sTweetedWords(nCtr)))

If (sTwitterTemp <> "") And (sHITsTemp <> "") Then 'Only records with content
If (sHITsTemp = sTwitterTemp) Then
file.WriteLine(sRecordIn)
nTweetRecs += 1
bFound = True
End If
End If
Next

'If here write records to output file, along with hits; but only if nTweetRecs <> 0.
If bFound = True Then
If nTweetRecs <> 0 Then
file.WriteLine(sRecordIn & Space(5) & "hits = " & Str(nTweetRecs))
End If
End If
Loop 'Read all HITS data
Else
MsgBox("Error opening " & sSourceHits)
Exit Sub 'Unable to open the source HITS data file
End If

'Close file
file.Close()
MsgBox("Data saved to " & sDestination)

Catch ex As Exception
MsgBox("Sequence2 has encountered an error and unable to continue.", ex.Message)
End Try
End Sub
```

Appendix I

PDF-Text-ToProcessedText

PDF-Text-ToProcessedText reads each line from the iSkysoft PDF converted Text file and writes out a new text file that includes corrections to formatting. The output is later used by other programs.

```
'*****************************************************************************************
'* Program: PDF-Text-ToProcessedText
'*    Date: 07-04-2020
'*      By: Robert W. Jones
'* Purpose: After PDFs are converted to text with iSkysoft, this program creates a
'*          text file to be used in other programs' processing of data. The iSkysoft
'*          text is in an unformatted layout.
'*****************************************************************************************
Public Class Form1

    Dim sArrayOfWords(65000) As String 'Array to hold words and check for dups
    Dim nArrayOfWordsCounter As Integer 'Keep track of how many words to check
    Private Sub Form1_Load(sender As Object, e As EventArgs) Handles MyBase.Load

'*****************************************************************************************
        '* Program: PDF-Text-ToProcessedText
        '*    Date: 07-04-2020
        '*      By: Robert W. Jones
        '* Purpose: After PDFs are converted to text with iSkysoft, this program creates a
        '*          text file to be used in other programs' processing of data. The iSkysoft
        '*          text is in an unformatted layout.

'*****************************************************************************************
Dim sSource As String = "" 'Source filename
Dim sSourceNew As String = "" 'Source from output, used as input
Dim sDestination As String = "" 'Destination filename
Dim sDestinationNoDups As String = "" 'No duplicates in final output
Dim sRecordOut As String = "" 'String record out
Dim sRecordIn As String = "" 'String record in
Dim sOut As String = "" 'String out
Dim nCtr As Integer = 0 'For loop counter
Dim nTotalWords As Integer = 0 'Record words processed

sSource = "C:\DISS-Work\IT-Artifact3-Experiment3\320-cr-00245.txt"
sDestination = "C:\DISS-Work\IT-Artifact3-Experiment3\320-cr-00245.out"
Dim file As System.IO.StreamWriter 'Outfile
file = My.Computer.FileSystem.OpenTextFileWriter(sDestination, False)

Try
If System.IO.File.Exists(sSource) = True Then
Dim objReader As New System.IO.StreamReader(sSource)
Do While objReader.Peek() <> -1
sRecordIn = objReader.ReadLine()
sOut = "" 'Reinitialize
```

Appendix I continued:

PDF-Text-ToProcessedText reads each line from the iSkysoft PDF converted Text file and writes out a new text file that includes corrections to formatting. The output is later used by other programs.

```
For nCtr = 1 To Len(RTrim(sRecordIn)) Step 1
'If not a space, build the word
If Mid(sRecordIn, nCtr, 1) <> Space(1) And Mid(sRecordIn, nCtr, 1) <> "." And Mid(sRecordIn, nCtr, 1)
<> "," Then
sOut = sOut & Mid(sRecordIn, nCtr, 1) 'This is how to build a good word.

Else
'If here, only write records that are over three characters long
If Len(RTrim(sOut)) >= 3 Then
'Track the word count for duplicate checking
If IsDuplicate(sOut) = False Then 'If word has not already been written, record it and write it.
'Track the word count for duplicate checking
nArrayOfWordsCounter = nArrayOfWordsCounter + 1
sArrayOfWords(nArrayOfWordsCounter) = sOut
file.WriteLine(sOut)
'Track the total words
nTotalWords = nTotalWords + 1
sOut = ""
End If
Else
sOut = ""
End If
End If
Next
Loop 'Read all input records from master text file
Else
MsgBox("Error opening " & sSource)
Exit Sub 'Unable to open the court data file
End If
 'Write remaining buffer, then close file
file.WriteLine(sOut)
file.Close()
Catch ex As Exception
MsgBox("Convert PDF text to processed words has encountered an error and unable to continue.")
End Try

MsgBox("Converted approximately " & RTrim(Str(nTotalWords)) & " total words to text.")

End Sub

Function IsDuplicate(sWord As String) As Boolean
Dim nDupCtr As Integer

IsDuplicate = False

For nDupCtr = 1 To nArrayOfWordsCounter Step 1
If LCase(sArrayOfWords(nDupCtr)) = LCase(sWord) Then
IsDuplicate = True
Exit For
```

Appendix I continued:

PDF-Text-ToProcessedText reads each line from the iSkysoft PDF converted Text file and writes out a new text file that includes corrections to formatting. The output is later used by other programs.

```
End If
Next
End Function
End Class
```

Appendix J

Court Case Negative Sentiments

abuse	explointing	leave	scare
along	exposed	liability	serious
arrested	fail	lies	sluggish
attacks	failed	limited	stinky
avoided	failing	loss	stopped
burden	fails	losses	suffer
complained	failure	lost	suffered
complaint	faulty	low	suspended
complaints	felony	lower	touted
conspiracy	fight	lowering	trouble
crime	fraud	misleading	unaware
criminal	gross	negative	uncertain
cut	grossly	no	unethical
dangerous	guilty	obstacles	unfair
death	hard	offend	victim
deceive	harm	offends	victims
deceived	harmed	offense	violate
deception	harsh	oppressive	violated
defeat	hide	pay	violation
delay	hoax	problem	violations
demand	illegal	problems	violence
denying	immoral	punish	weapon
devastating	imposed	questioned	wrong
difficult	injured	reckless	
dispute	injury	refuse	
disregard	injured	refused	
error	injury	risk	
excluded	lawsuit	risks	

Appendix K

Descriptions for Negative Sentiments

NRC-Hash-Sent-negScore – Words are annotated according to eight emotions: joy, trust, sadness, anger, surprise, fear, anticipation, and disgust, and two polarity classes: positive and negative. There are many words that are not as- sociated with any emotional state and are tagged as neutral (Bravo-Marquez, Frank, & Pfahringer, 2015).

SentiWordnet-negScore – A lexical resource explicitly devised for supporting sentiment classification and opinion mining (Baccianella, Esuli, & Sebastiani, 2010) that relates to the highest top ten ranked negative synsets.

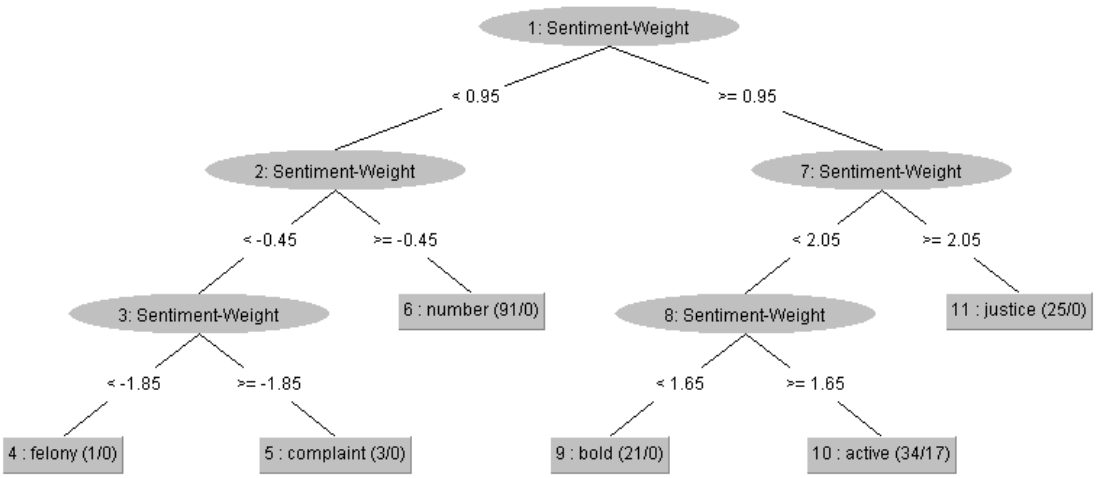
AFINN-negScore – Negative words scored from -1 to -5, includes slang, obscene words, acronyms and Web jargon (Bravo-Marquez, Frank, & Pfahringer, 2015).

S140-negScore – According to Bandhakavi, Wiratunga, and Massie (2018), emotion-aware polarity lexicons for Twitter Sentiment Analysis uses Twitter API, to include a collection of 177 negative manually annotated tweets, yet includes a collection to 1.6 million sourced tweets.

While other negative sentiment attributes exist within the WEKA package AffectiveTweets, the antecedent was the only attributes carrying the lowest values, and remained consistent throughout all lab experiments.

Appendix L

Random Tree Illustration from Experiment 1



The root sentiment weight has two possible initial paths. For the experiment, values less than 0.95 are first noted. This leads to the left of the tree with possible paths of less than -0.45 and values greater. Focus is on the left branch and leads to lexicon weights of less than -1.85 or greater than or equal to this value. Both ending nodes carry significance. The first being the class of *felony* 1/0 and *complaints* 3/0. This means both nodes' data was discovered within tweets that might correlate to fraud. This is important at an organizational level within the preemployment vetting, just as negative scoring carries significant value.

Appendix M

Kappa Statistics

Kappa is widely accepted in the field of content analysis (Carletta, 1996) and is used to assess the agreement or reliability between two observers who are performing a test which has a categorical variable (McLintic, 2009). According to Sahoo (2013), classifiers provide greater accuracy when Kappa statistic is greater than zero. According to McHugh (2012), the interpretation of Cohen's (1960) Kappa development suggests the following Kappa results. This study's lab experiments found the Kappa statistics in the range for substantial, or a nearly perfect agreement ranging from 0.763 to 1.0.

0 = No agreement

0.01–0.20 = Slight agreement

0.21–0.40 = Fair agreement

0.41– 0.60 = Moderate agreement

0.61–0.80 = Substantial agreement

0.81–1.00 = Nearly a perfect agreement

Appendix N

Dr. Stacie Petter Permission

From: Petter, Stacie
Sent: Thursday, October 15, 2020 10:36:14 AM
To: Robert Jones <rj631@mynsu.nova.edu>
Cc: GERARD DE LEOZ
Subject: Re: Permission

NSU Security WARNING: This is an external email. Do not click links or open attachments unless you recognize the sender and know that the content is safe.

Bob,

I'm glad you found our paper and figure useful in our *European Journal of Information Systems*. The journal retains copyright, but usually, for something like a dissertation, it is acceptable to include the figure, but then cite the source of the figure.

It can sometimes be more complicated if you want to publish a journal or conference article with the figure depending on the publisher's policies. Sometimes you have to get permission from the publisher to include content that was published elsewhere.

I hope that helps.

Best,
Stacie

Figure N1. Dr. Stacie Petter Permission

Appendix O

Dr. Udo Bub Permission

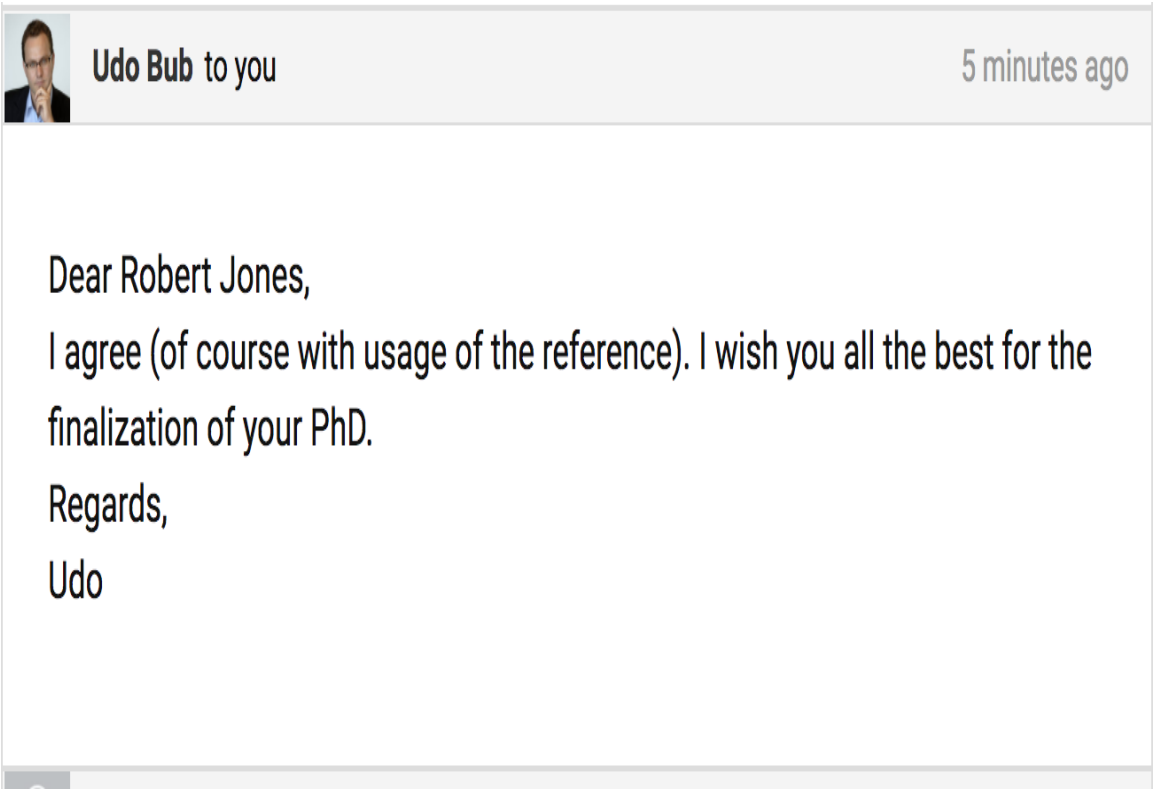
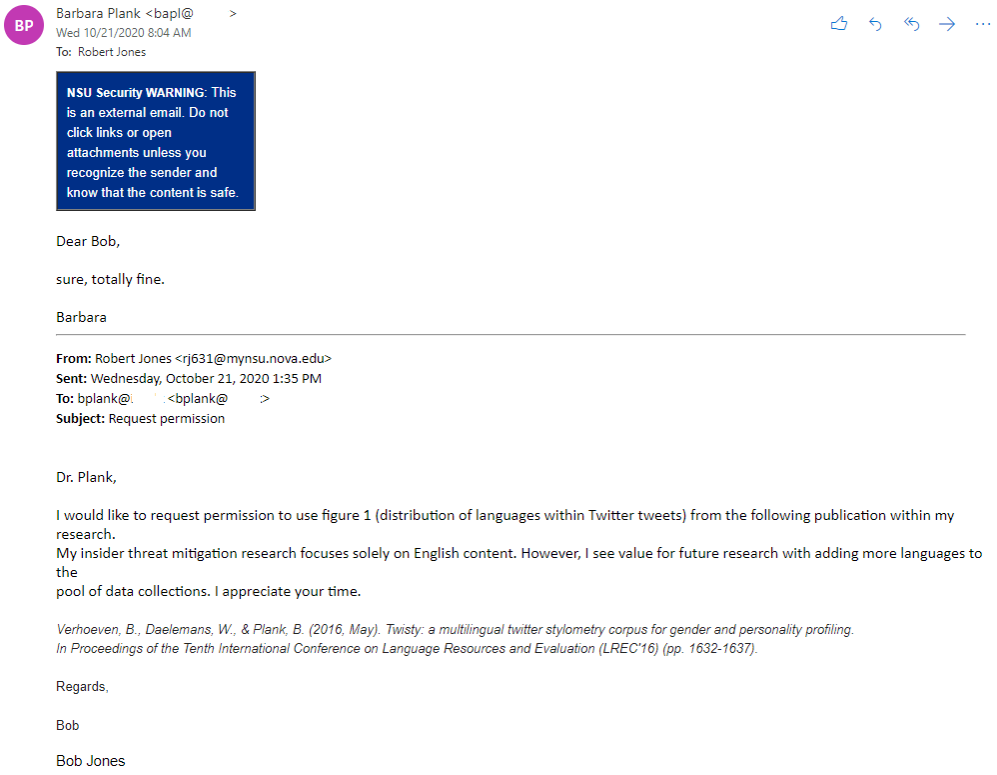


Figure O1. Dr. Udo Bub Permission.

Appendix P

Dr. Barbara Plank Permission



BP

Barbara Plank <bapl@ >
Wed 10/21/2020 8:04 AM
To: Robert Jones

NSU Security WARNING: This is an external email. Do not click links or open attachments unless you recognize the sender and know that the content is safe.

Dear Bob,

sure, totally fine.

Barbara

From: Robert Jones <rj631@mynsu.nova.edu>
Sent: Wednesday, October 21, 2020 1:35 PM
To: bplank@. .<bplank@ :>
Subject: Request permission

Dr. Plank,

I would like to request permission to use figure 1 (distribution of languages within Twitter tweets) from the following publication within my research.
My insider threat mitigation research focuses solely on English content. However, I see value for future research with adding more languages to the
pool of data collections. I appreciate your time.

Verhoeven, B., Daelemans, W., & Plank, B. (2016, May). Twisty: a multilingual twitter stylometry corpus for gender and personality profiling. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 1632-1637).

Regards,

Bob

Bob Jones

Figure P1. Dr. Barbara Plank Permission

Appendix Q

Experiment 1 TPR and FPR Results

TPR (Zero Value) Classes and Associated Negative Lexicons

complaint (-1.2)	*felony (-2.5)*

TPR (Greater than 0 and less than 1Value) Classes and Associated Negative Lexicons

active	parties

TPR (1Value) Classes and Associated Negative Lexicons

bold	justice	number

Appendix R

Experiment 2 TPR and FPR Results

TPR (Zero Value) Classes and Associated Negative Lexicons

assets	credits	*fraud (-2.8)*
liability (-0.8)	promises	substantial

TPR (Greater than 0 and less than 1Value) Classes and Associated Negative Lexicons

creating	shares

TPR (1Value) Classes and Associated Negative Lexicons

created	*hide (-0.7)*	legal
limited (-0.9)	matter	united

Appendix S

Experiment 3 TPR and FPR Results

TPR (Zero Value) Classes and Associated Negative Lexicons

acceptance	approval	assets
credits	determination	determined
felony (-2.5)	gains	granting
imposed (-0.3)	improvement	*liability (-0.8)*
losses (-1.7)	mandatory	*offenses (-1.5)*
promises	*questioned (-0.4)*	recommended
secured	*suspended (-2.1)*	*victims (-1.3)*
violation (-2.2)	*violence (-3.1)*	*weapon (-1.2)*
wells		

TPR (Greater than 0 and less than 1Value) Classes and Associated Negative Lexicons

arrested (-2.1)	*conspiracy (-2.4)*	engaged
entitled	*error (-1.7)*	*victim (-1.1)*

TPR (1Value) Classes and Associated Negative Lexicons

ability	*abuse (-3.2)*	accept
accepted	accepting	agreement
allow	approved	benefit
commit	committed	credit
crime (-2.5)	*criminal (-2.4)*	*dangerous (-2.1)*
dear	*death (-2.9)*	defense
effective	fine	fit
fraud (-2.8)	grant	*gross (-2.1)*
guilty (-1.8)	honorable	*injury (-1.8)*
interest	justice	*leave (-0.2)*
loss (-1.3)	*low (-1.1)*	matter
number	*offense (-1.0)*	original
outstanding	parties	*pay (-0.4)*
please	*risk (-1.1)*	safety
secure	sentence	smart
special	united	

Appendix T

Experiment 4 TPR and FPR Results

TPR (Zero Value) Classes and Associated Negative Lexicons

admits	advanced	approval
asset	assured	awarded
benefits	*cancer (-3.4)*	charities
commitment	*complained (-1.7)*	*complaint (-1.2)*
complaints (-1.7)	*deceive (-1.7)*	*deceived (-1.9)*
delay (-1.3)	*denying (-1.4)*	effectively
engaged	ensure	entitled
excluded (-1.4)	extends	*fail (-2.5)*
failing (-2.3)	*fails (-1.8)*	*faulty (-1.3)*
greater	increase	increased
lawsuit (-0.9)	*misleading (-1.7)*	*offends (-2.0)*
oppressive (-1.7)	profits	promote
questioned (-0.4)	recommended	recommends
refused (-1.2)	responsible	safest
safety	satisfied	saved
substantial	*suffer (-2.5)*	*suffered (-2.2)*
trusting	*unaware (-0.8)*	*unethical (-2.3)*
unfair (-2.1)	*violate (-2.2)*	*violation (-2.2)*
violations (-2.4)	*warn (-0.4)*	

TPR (Greater than 0 and less than 1Value) Classes and Associated Negative Lexicons

acceptable	agree	*attacks (-1.9)*
award	care	certain
challenges	clear	committed
creates	*cut (-1.1)*	desire
devastating (-3.3)	dream	engage
engaging	exclusive	*exposed (-0.3)*
failed (-2.3)	fair	giving
guarantee	*hard (-0.4)*	*harm (-2.5)*
honest	*immoral (-2.0)*	important
injured (-1.7)	*injury (-1.8)*	interest
interests	legal	*lies (-1.8)*
limited (-0.9)	*lost (-1.3)*	*low (-1.1)*
lower (-1.2)	matter	matters
no (-1.2)	parties	party
please	*problems (-1.7)*	protect
punish (-2.4)	*refuse (-1.2)*	relief
respect	*risk (-1.1)*	*risks (-1.1)*
safe	*scare (-2.2)*	*serious (-0.3)*
share	significant	strong

Appendix T continued:

Experiment 4 TPR and FPR Results

TPR (Greater than 0 and less than 1Value) Classes and Associated Negative Lexicons

superior	Support	sure
top	*trouble (-1.7)*	truth
united	Value	want
well	*wrong (-2.1)*	yes

TPR (1Value) Classes and Associated Negative Lexicons

admitted	Best	*demand (-0.5)*
fight (-1.6)	*fraud (-2.8)*	growing
illegal (-2.6)	*lowering (-1.0)*	*touted (-0.2)*

Appendix U

Experiment 5 TPR and FPR Results

TPR (Zero Value) Classes and Associated Negative Lexicons

	active	actively
approval	*avoided (-1.4)*	awarded
benefit	boosted	*burdens (-1.5)*
challenges	cleaner	*complaint (-1.2)*
deceive (-1.7)	*deceived (-1.9)*	*deception (-1.9)*
delay (-1.3)	determination	determined
dispute (-1.7)	*disregard (-1.1)*	efficient
engaging	ensure	*excluded (-1.4)*
exploiting (-1.9)	*failures (-2.0)*	*futile (-1.9)*
grossly (-0.9)	*harmed (-2.1)*	*harsh (-1.9)*
hoax (-1.1)	*immoral (-2.0)*	improvements
inability (-1.7)	increased	innovative
liability (-0.8)	*losses (-1.7)*	*misleading (-1.7)*
offend (-1.2)	parties	profit
profits	promise	protects
punish (-2.4)	*reckless (-1.7)*	relief
respective	responsible	satisfied
satisfy	shared	*sluggish (-1.7)*
sophisticated	*stinky (-1.5)*	substantial
touted (-0.2)	trusted	*uncertain (-1.2)*
unethical (-2.3)	*unfair (-2.1)*	*unjust (-2.3)*
value	*victims (-1.3)*	*violated (-2.4)*
violation (-2.2)	*violations (-2.4)*	virtue
vision		

TPR (Greater than 0 and less than 1Value) Classes and Associated Negative Lexicons

ability	accomplish	active
actively	admit	authority
award	benefits	better
burden (-1.9)	care	certain
clean	clear	committed
confidence	*confusion (-1.2)*	create
created	creates	*defeat (-2.0)*
defense	*delay (-1.3)*	*difficult (-1.5)*
effective	engaged	entitled
escape	exclusive	*failed (-2.3)*
failure (-2.3)	fair	fit
fraud (-2.8)	free	friendly
giving	good	

Appendix U continued:

Experiment 5 TPR and FPR Results

TPR (1Value) Classes and Associated Negative Lexicons

admitted	*alone* (-1.0)	best
great	*illegal (-2.6)*	*injury (-1.8)*
negative (-2.7)	*pay* (-0.4)	successful

Appendix V

Experiment 6 TPR and FPR Results

TPR (Zero Value) Classes and Associated Negative Lexicons

abuse (-3.2)	admits	admitted
asset	assurances	*avoided (-1.4)*
committing	compelled	competition
complained (-1.7)	*complaint (-1.2)*	consent
demanded (-0.9)	*demanding (-0.9)*	*destruction (-2.7)*
determined	*disregard (-1.1)*	efficient
engaged	engagement	ensuring
entitled	extends	favor
favors	*hacked (-1.7)*	*harmed (-2.1)*
integrity	intellectual	merits
preventing (-0.1)	profit	profits
promises	*refused (-1.2)*	*refusing (-1.7)*
respectfully	*restricting (-1.6)*	secured
substantial	*suspected (-0.9)*	*threatened (-2.0)*
threatens (-1.6)	*threats (-1.8)*	*unacceptable (-2.0)*
unethical (-2.3)	*unjust (-2.3)*	*violate (-2.2)*
violated (-2.4)	*violation (-2.2)*	virtue

TPR (Greater than 0 and less than 1Value) Classes and Associated Negative Lexicons

advantage	authority	benefit
benefits	*destroying (-2.6)*	enjoyed
failing (-2.3)	*failure (-2.3)*	greater
lawsuit (-0.9)	protected	relief
secure	*steal (-2.2)*	*suffered (-2.2)*
threatening (-2.4)	united	valuable
worthy		

TPR (1Value) Classes and Associated Negative Lexicons

agree	agreed	agreement
alone (-1.0)	award	certain
clear	committed	created
credit	*crime (-2.5)*	*criminal (-2.4)*
damage (-2.2)	*demand (-0.5)*	*destroy (-2.5)*
difficult (-1.5)	easily	excuse
failed (-2.3)	*fraud (-2.8)*	giving
great	*gross (-2.1)*	*harm (-2.5)*
illegal (-2.6)	interest	legal

Appendix V continued:

Experiment 6 TPR and FPR Results

TPR (1Value) Classes and Associated Negative Lexicons

lies (-1.8)	*limited (-0.9)*	*loss (-1.3)*
lost (-1.3)	*low (-1.1)*	matter
number	original	parties
party	please	prevent
promise	promised	protect
respect	security	share
steal (-2.2)	*stop (-1.2)*	success
suffer (-2.5)	Support	*threat (-2.4)*
value	Well	

Appendix W

Summary from Experiments

Experiment	TPR	FPR	Precision	Kappa	Classification	Theory	Negative Lexicons
1	0.873	0.039	0.539	0.8055	Trees Random Forest	RAT	2
2	0.550	0.009	0.999	0.8106	Trees Random Forest	RAT / TPB	4
3	0.948	0.003	0.958	0.9449	Discriminative Multinomial	RAT / TPB	30
4	0.700	0.005	0.736	0.6822	Naive Bayes	TRA / RAT	58
5	0.780	0.004	0.995	0.7588	Naive Bayes	PMT / RAT	48
6	0.929	0.002	0.492	0.9257	Discriminative Multinomial	RAT	43

Appendix X

Behavioral Theories within Lab Experiments

Experiment	Theory
1	Routine Activity Theory
2	Routine Activity Theory / Theory of Planned Behavior
3	Routine Activity Theory / Theory of Planned Behavior
4	Routine Activity Theory / Theory of Reasoned Action
5	Routine Activity Theory / Protection Motivation Theory
6	Routine Activity Theory

Appendix Y

Behavioral Theory Correlations within Lab Experiments

Experiment	Theory	Correlation
1	Routine Activity Theory	The turn of unfortunate events shared by D'Addona (2019) aligns with the RAT. In this instance, the offender leveraged his tenure and promotions through the ranks to place himself in a position with access to financial components within the organization. According to Cohen & Felson's (1979), definition of RAT. The circumstances surrounding the embezzlement was demonstrated through the lack of capable guardians against criminal activities, and the offender sought suitable targets; all in alignment with the theory.
2	Routine Activity Theory / Theory of Planned Behavior	After reviewing the court's news release and reviewing the case as filed with the courts, this particular case appeared to align with two behavioral theories: the RAT and the TPB. As Cohen and Felson (1979) posited, unlawful activities are brought together through conditions exhibited in this case, along with investors (the targets) and lacked protectors to these types of criminal activities. Equally, the TPB demonstrates the insufficiency of following any type of behavioral control, antecedents of attitudes, subjective norms, and perceived behavioral control that lead to predictors with intentions and actions (Ajzen, 1991).
3	Routine Activity Theory / Theory of Planned Behavior	Review of the court's case appeared to align with two behavioral theories: the RAT and the TPB. As in the previous experiment and applicable is Cohen and Felson's (1979) theory with unlawful activities coming together through conditions exhibited in this case; along with banks and credit unions (the targets) and lacked protectors to these types of criminal activities. Equally important, the TPB demonstrates the insufficiency of following any type of behavioral control, antecedents of attitudes, subjective norms, and perceived behavioral control that lead to predictors with intentions and actions (Ajzen, 1991), and appears to be demonstrated by greed. The offenders appeared to target investors who were all blind-eyed to activities outside of their knowledge using an elaborate strategy of being in control of schemes that lacked checks and balances.

Appendix Y continued:

Behavioral Theory Correlations within Lab Experiments

Experiment	Theory	Correlation
4	Routine Activity Theory / Theory of Reasoned Action	The impact of Artifact 4 included both business partners and consumers and appeared to fall within two theories. The theory of reasoned action (TRA) is used to reveal the meaningful effects of attitudes and subjective norms. In this particular case, the business appeared to follow a subjective norm common with other business owners' practices. The behavior is in alignment with Hale, Householder and Greene's (2002) assessment to subjective norms. Similarly, the business specifically sought suitable targets in the absence of guardians against crime (Cohen & Felson, 1979) in order to carry out their business practices and part of the routine activity theory (RAT).
5	Routine Activity Theory / Protection Motivation Theory	The theories that correlate best with this case are protection motivation theory (PMT) and routine activity theory (RAT). Protection motivation theory represents the cognitive processes to mediate the persuasive effects of a fear appeal by arousing protection motivation. In this case, it appears dwindling sales was a motivation and according to Maddux and Rogers (1983), the protection motivation came from self-preservation with keeping the business afloat. Furthermore, one could theorize the danger felt by the manufacturer might be construed with the fear from competitors and led to the business finding suitable targets; the consumer and in alignment with RAT.
6	Routine Activity Theory	Similarly with other experiments' interpretation with behavioral theories, RAT best applies to this case. In this instance, the contractor appeared to knowingly select a target thought to be incapable of defending itself, which revealed the absence of capable guardians against crime (Cohen & Felson, 1979). However, what the contractor did not realize at the time was the bank's trade secrets on the laptop and the willingness to prosecute based on theft.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211.

Alahmadi, B. A., Legg, P. A., & Nurse, J. R. C. (2015). Using Internet activity profiling for insider-threat detection. *Institute for Systems and Technologies of Information, Control and Communication*.

Alshboul, Y., & Streff, K. (2017). Beyond cybersecurity awareness: Antecedents and satisfaction. In *Proceedings of the 2017 International Conference on Software and e-Business,* 85-91.

Amazon.com (2021). Amazon Comprehend Features. Retrieved from https://aws.amazon.com/comprehend/features/

Ambre, A., & Shekokar, N. (2015). Insider threat detection using log analysis and event correlation. *Procedia Computer Science*, *45*, 436-445.

Amin, M. N., & Habib, M. A. (2015). Comparison of different classification techniques using WEKA for hematological data. *American Journal of Engineering Research, 4*(3), 55-61.

Anderson, D. (2020). United States Attorney. U.S. District Court. Northern District of California, San Francisco.

Antoniou, G., Billington, D., & Maher, M. J. (1999). On the analysis of regulations using defeasible rules. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences.*

Azar, A. T., Elshazly, H. I., Hassanien, A. E., & Elkorany, A. M. (2014). A Random Forest Classifier for Lymph Diseases. *Computer Methods and Programs in Biomedicine*, *113*(2), 465-473.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordnet 3.0: an enhanced lexical

BaMaung, D., McIlhatton, D., MacDonald, M., & Beattie, R. (2018). The enemy within? The connection between insider threat and terrorism. *Studies in Conflict & Terrorism*, *41*(2), 133-150.

Bandhakavi, A., Wiratunga, N., & Massie, S. (2018). Emotion-aware polarity lexicons for Twitter sentiment analysis. *Expert Systems*, e12332.

Bauer, S., Bernroider, E. W., & Chudzikowski, K. (2017). Prevention is better than cure! Designing information security awareness programs to overcome users' non-compliance with information security policies in banks. *Computers & Security, 68*, 145-159.

Bell, A. J., Rogers, M. B., & Pearce, J. M. (2019). The insider threat: Behavioral indicators and factors influencing likelihood of intervention. *International Journal of Critical Infrastructure Protection*, *24*, 166-176.

Benferhat, S., Boudjelida, A., Tabia, K., & Drias, H. (2013). An intrusion detection and alert correlation approach based on revising probabilistic classifiers using expert knowledge. *Applied Intelligence*, *38*(4), 520-540.

Berski, A. (2016). Internal disciplinary procedures–internet and social media. *Dilemmas of Bilateral Relations*.

BigML.com (2021). Machine Learning made beautifully simple for everyone. Retrieved from https://bigml.com

Biswas, B., Mukhopadhyay, A., & Gupta, G. (2018). Leadership in action: How top hackers behave: A big-data approach with text-mining and sentiment analysis. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, 1752-1761.

Bonta, V., & Janardhan, N. K. N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1-6.

Branz, L., & Brockmann, P. (2018). Sentiment Analysis of Twitter Data: Towards Filtering, Analyzing and Interpreting Social Network Data. In *Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems*, 238-241.

Bravo-Marquez, F., Frank, E., & Pfahringer, B. (2015). Positive, negative, or neutral: Learning an expanded opinion lexicon from emoticon-annotated tweets. *In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015*, 1229-1235.

Bravo-Marquez, F., Frank, E., Pfahringer, B., & Mohammad, S. M. (2019). AffectiveTweets: a WEKA package for analyzing affect in tweets.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Calculator.net (2020). Random Number Generator. *Comprehensive Version.* Retrieved from https://www.calculator.net/random-number-generator.html

Cappelli, D., Moore, A., Trzeciak, R., & Shimeall, T. J. (2009). Common sense guide to prevention and detection of insider threats. *CERT*.

Carletta, J. (1996). Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics*, *22*(2), 249-254.

Carpenter, M. L. (2017). Social network usage in the hiring process. *WRIT: GSW Journal of First-Year Writing*, *1*(2), 9.

Catrantzos, N. (2018). Insider threat: Applying no dark corners defenses. *Handbook of Security Science*, 1-20.

Chen, K., Wang, P., Lee, Y., Wang, X., Zhang, N., Huang, H., ... & Liu, P. (2015). Finding unknown malice in 10 seconds: Mass vetting for new threats at the Google Play scale. In *24th {USENIX} Security Symposium. {USENIX} Security 15*, 659-674.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 1-13.

Chinchani, R., Iyer, A., Ngo, H. Q., & Upadhyaya, S. (2005). Towards a theory of insider threat assessment. In *2005 International Conference on Dependable Systems and Networks (DSN'05)*, 108-117.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 588-608.

Cole, E. (2015). Insider threats and the need for fast and directed response. *SANS Institute InfoSec Reading Room, Tech. Rep*.

Costa, D. L., Albrethsen, M. J., & Collins, M. L. (2016). *Insider threat indicator ontology* (No. CMU/SEI-2016-TR-007). Carnegie-Mellon Univ, Pittsburgh, PA.

Cram, W. A., Proudfoot, J. G., & D'Arcy, J. (2017). Organizational information security policies: a review and research framework. *European Journal of Information Systems*, *26*(6), 605-641.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 124-130.

Datawrapper.de (2021). Enrich your stories with charts, maps, and tables. Retrieved from https://www.datawrapper.de

DataRobot.com (2021). Overcoming Hurdles and Achieving Success with AI. DataRobot is the leading end-to-end enterprise AI platform that automates and accelerates every step of your path from data to value. Retrieved from https://www.datarobot.com

D'Addona, D. (2019). Swimming World Magazine. Online home of the International Swimming Hall of Fame. *John Bitter, Former Santa Clara Coach Arraigned on Embezzlement Charges.* Retrieved from https://www.swimmingworldmagazine.com/news/john-bitter-former-santa-clara-coach-arraigned-on-embezzlement-charges/

De Leoz, G., & Petter, S. (2018). Considering the social impacts of artefacts in information systems design science research. *European Journal of Information Systems*, *27*(2), 154-170.

Duarte, N., Llanso, E., & Loup, A. C. (2018). Mixed Messages? The Limits of Automated Social Media Content Analysis. In *FAT*, 106.

Delarosa, J. (2015). From due diligence to discrimination: Employer use of social media vetting in the hiring process and potential liabilities. *Loyola of Los Angeles Entertainment Law Review*, *35*(3), 249.

Egieyeh, S., Syce, J., Malan, S. F., & Christoffels, A. (2018). Predictive classifier models built from natural products with antimalarial bioactivity using machine learning approach. *PloS one*, *13*(9), e0204644.

Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, *8*(1), 19-20.

Fick, M., & Dave, P. (2019). Facebook's flood of languages leave it struggling to monitor content. Retrieved from https://www.reuters.com/article/us-facebook-languages-insight/facebooks-flood-of-languages-leave-it-struggling-to-monitor-content-iduskcn1rz0dw

Fischbacher-Smith, D. (2015). The enemy has passed through the gate: Insider threats, the dark triad, and the challenges around security. *Journal of Organizational Effectiveness: People and Performance*, *2*(2), 134-156.

Flowerday, S. V., & Tuyikeze, T. (2016). Information security policy development and implementation: The what, how and who. *Computers & Security*, *61*, 169-183.

FoxBusiness.com (2014). California man pleads guilty in nationwide auto loan scam before Pittsburgh federal judge. *Associated Press.* Lifestyle and Budget.

Gallagher, C., McMenemy, D., & Poulter, A. (2015). Management of acceptable use of computing facilities in the public library: Avoiding a panoptic gaze? *Journal of Documentation*, *71*(3), 572-590.

Garner, S. R. (1995). Weka: The Waikato environment for knowledge analysis. In *Proceedings of the New Zealand Computer Science Research Students Conference*, *1995*, 57-64.

GitHub.com (2020). FacePager. Retrieved from https://github.com/strohne/Facepager/blob/master/README.md

Glackin, C., & Bible, L. (2019). *Alarming issues: The case of the corrupt security industry executive*. Sage Business Cases Originals.

Gritzalis, D., Stavrou, V., Kandias, M., & Stergiopoulos, G. (2014). Insider threat: Enhancing BPM through social media. In *2014 6th International Conference on New Technologies, Mobility and Security (NTMS).* 1-6.

Hale, J. L., Householder, B. J., & Greene, K. L. (2002). The theory of reasoned action. *The persuasion handbook: Developments in Theory and Practice*, *14*, 259-286.

HG.org (2020). H.G. Legal Resources. Retrieved from https://www.hg.org/legal-articles/most-common-types-of-white-collar-crimes-40448

Hielscher, E., & Waghorn, G. (2015). Managing disclosure of personal information: An opportunity to enhance supported employment. *Psychiatric Rehabilitation Journal*, *38*(4), 306.

Hunker, J., & Probst, C. W. (2011). Insiders and Insider Threats-An Overview of Definitions and Mitigation Techniques. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 2*(1), 4-27.

Hubballi, N., & Suryanarayanan, V. (2014). False alarm minimization techniques in signature-based intrusion detection systems: A survey. *Computer Communications*, *49*, 1-17.

Hutto, C. J., & Gilbert, E. (2014a). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 216-225.

Hutto, C.J., & Gilbert, E. (2014b). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of *Eighth International Conference on Weblogs and Social Media. 81*, 82.

Intelligence and National Security Alliance (2015). Explanation of INSA-Developed Insider Threat Definition. Retrieved from https://www.insaonline.org/explanation-of-insa-insider-threat-definition/

Ismail, W. B. W., & Yusof, M. (2018). Mitigation strategies for unintentional insider threats on information leaks. *International Journal of Security and Its Applications*, *12*(1), 37-46.

Jaafari, J., & Lewis, N. (2019). In court, where are Siri and Alexa? When it comes to setting the record straight, court reporting technology is still not up to speed. *The Marshall Project Justice Lab*. Retrieved from https://www.themarshallproject.org/2019/02/14/in-court-where-are-siri-and-alexa

Jarrett, L., & Borger, G. (2017). Obama commutes sentence of Chelsea Manning. *CNN Politics*. Retrieved from http://www.cnn.com/2017/01/17/politics/chelsea-manning-sentence-commuted.

Jeske, D., & Holland, P. (2019). Employer and Employee Vetting: Reputation

Jeske, D., Lippke, S., & Shultz, K. S. (2019). Predicting self-disclosure in recruitment in the context of social media screening. *Employee Responsibilities and Rights Journal*, 1-14.

Jünger, J., & Keyling, T. (2013). Facepager. An application for generic data retrieval through APIs. Retrieved from https://github. com/strohne/Facepager.

Justice.gov (2020). United States Department of Justice. The United States Attorney's Office. Northern District of California. Retrieved from https://www.justice.gov/usao-ndca/pr/san-francisco-venture-capitalist-charged-wide-ranging-schemes-defraud

Justice.gov (2020). United States Department of Justice. The United States Attorney's Office. Northern District of California. Retrieved from: https://www.justice.gov/usao-ndca/pr/san-francisco-venture-capitalist-charged-wide-ranging-schemes-defraud

Kamburugamuve, S., Wickramasinghe, P., Ekanayake, S., & Fox, G. C. (2018). Anatomy of machine learning algorithm implementations in MPI, Spark, and Flink. *The International Journal of High-Performance Computing Applications*, *32*(1), 61-73.

Kandias, M., Stavrou, V., Bozovic, N., & Gritzalis, D. (2013). Proactive insider threat detection through social media: The YouTube case. In *Proceedings of the 12th ACM workshop on Workshop on Privacy in the Electronic Society*, 261-266.

Kauh, J., Lim, W., Kwon, K., Lee, J. E., Kim, J. J., Ryu, M., & Cha, S. H. (2017). Indicator-based behavior ontology for detecting insider threats in network systems. *KSII Transactions on Internet & Information Systems*, *11*(10), 5062-5079.

Kaur, J., & Saini, J. R. (2015). A Study of Text Classification Natural Language Processing Algorithms for Indian Languages. *The VNSGU Journal of Science Technology*, *4*(1), 162-167.

Kenazag, Tayeb, Mahdi, & Aiash (2016). Toward an efficient ontology-based event correlation in SIEM., 139-146.

Kotthoff, L., Thornton, C., & Hutter, F. (2017). User guide for auto-WEKA version 2.6. *Dept. Computer Science, University British Columbia, BETA lab, Vancouver, BC, Canada, Tech. Rep*, *2*.

Kotthoff, L., Thornton, C., Hoos, H. H., Hutter, F., & Leyton-Brown, K. (2017). Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. *The Journal of Machine Learning Research*, *18*(1), 826-830.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221-232.

Krull, K. E. (2016). *The threat among us: Insiders intensify aviation terrorism* (No. PNNL-25689). Pacific Northwest National Lab., Richland, WA.

Kühn, S., & Nieman, A. (2017). Can security vetting be extended to include the detection of financial misconduct? *African Security Review*, *26*(4), 413-433.

Lee, S., & Huh, J. H. (2019). An effective security measures for nuclear power plant using big data analysis approach. *The Journal of Supercomputing*, *75*(8), 4267-4294.

Legg, P. A. (2015). Visualizing the insider threat: challenges and tools for identifying malicious user activity. In *Visualization for Cyber Security (VizSec), 2015 IEEE Symposium*, 1-7.

Leonard, L. N., Cronan, T. P., & Kreie, J. (2004). What influences IT ethical behavior intentions—planned behavior, reasoned action, perceived importance, or individual characteristics? *Information & Management*, *42*(1), 143-158.

Linkov, I., Poinsatte-Jones, K., Trump, B. D., Ganin, A. A., & Kepner, J. (2019). Rulemaking for insider threat mitigation. In *Cyber resilience of systems and networks*, *Springer*, 265-286.

Linton, D. (2019). Amicus reporting. *Virginia Court Reporters: Ashburn, Dulles, Fairfax, Loudoun County, Reston, and all of Northern Virginia*. Retrieved from https://www.amicusreporting.org

Liu, L., De Vel, O., Han, Q. L., Zhang, J., & Xiang, Y. (2018). Detecting and preventing cyber insider threats: A survey. *IEEE Communications Surveys & Tutorials*, *20*(2), 1397-1417.

Loffi, J. M., & Wallace, R. J. (2014). The unmitigated insider threat to aviation (Part 1): a qualitative analysis of risks. *Journal of Transportation Security*, *7*(4), 289-305.

Lomas, D. W. (2019). "Crocodiles in the corridors": Security vetting, race and Whitehall, 1945–1968. *The Journal of Imperial and Commonwealth History*, 1-30.

Lupton, D., & Michael, M. (2017). "Depends on who's got the data": Public understandings of personal digital dataveillance. *Surveillance & Society*, *15*(2), 254-268.

Maasberg, M., Warren, J., & Beebe, N. L. (2015). The dark side of the insider: Detecting the insider threat through examination of dark triad personality traits. In *48th Hawaii International Conference on System Sciences,* 3518-3526.

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of experimental social psychology*, *19*(5), 469-479.

Management Challenges in the Information Age. *Contemporary HRM Issues in the 21st Century*, *Emerald Publishing Limited*, 149-157.

Manning, C. D., Schütze, H., & Raghavan, P. (2008). *Introduction to information retrieval*. Cambridge University Press. Retrieved from https://nlp.stanford.edu/IR-book/pdf/15svm.pdf

Martin, P. W. (2008). Online Access to Court Records-from Documents to Data, Particulars to Patterns. *Vill. L. Rev.*, *53*, 855.

McHugh Mary, L. (2012). Interrater Reliability: The Kappa Statistic. Biochem Med (Zagreb). U.S. National Library of Medicine. National Institutes of Health. 22(3), 276-282. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

McLintic, A. (2009). Understanding Statistics.

Mehra, G. (2017). 105 Leading social networks worldwide. *Practical eCommerce*. Retrieved from https://www.practicalecommerce.com/105-leading-social-networks-worldwide.

Merriam-Webster.com (2020). Definition of lexicon. Retrieved from https://www.merriam-webster.com/dictionary/lexicon

Mitrou, L., Kandias, M., Stavrou, V., & Gritzalis, D. (2014). Social media profiling: A panopticon or omniopticon tool? In *Proceedings of the 6th Conference of the Surveillance Studies Network,* 1-15.

MLbase.org (2021). Distributed Learning Made Easy. Retrieved from http://mlbase.org

Mohammad, S. M., & Turney, P. D. (2013). NRC emotion lexicon. *National Research*

Nawawi, A., & Salin, A. S. A. P. (2018). Employee fraud and misconduct: empirical evidence from a telecommunication company. *Information & Computer Security*, *26*(1), 129-144.

Offermann, P., Levina, O., Schönherr, M., & Bub, U. (2009). Outline of a design science research process. In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology,* 1-11.

Oladimeji, T. O., Ayo, C. K., & Adewumi, S. E. (2019). Review on insider threat detection techniques. *Journal of Physics: Conference Series, 1299*(1), 12046.

PACER.GOV (2020). Public Access to Court Electronic Records. Retrieved from https://www.pacer.gov

Panda, M. (2018). Developing an Efficient Text Pre-Processing Method with Sparse Generative Naive Bayes for Text Mining. *International Journal of Modern Education & Computer Science, 10*(9).

Park, R. K., Lim, J. I., Kwon, H. Y., & Choi, J. Y. (2017). A Study on Korea's Information security management system: An insider threat perspective. The steering committee of the World Congress in Computer Science, Computer Engineering and Applied Computing, In *Proceedings of the International Conference on Security and Management,* 61-67.

Park, W., You, Y., & Lee, K. (2018). Detecting potential insider threat: Analyzing insiders' sentiment exposed in social media. *Security and Communication Networks, 2018*.

Patil, M. S., Kamdar, J. K., & Khatri, (2014). C. B. BIG DATA–An Overview, *International Journal of Engineering Research & Technology (IJERT), 3*(7).

Patil, S. S., & Sonavane, S. P. (2017). Improved classification of large imbalanced data sets using rationalized technique: Updated Class Purity Maximization Over Sampling Technique (UCPMOT). *Journal of Big Data, 4*(1), 49.

Paxata.com (2021). The Data Prep for AI Toolkit. Retrieved from https://www.paxata.com

Pelton, V. J. (2017). The Enemy Among Us: The Insider Threat. *J. Air L. & Com., 82*, 519.

Powers, N. (2017). What history teaches us about today's insider threats. Retrieved from https://deltarisk.com/blog/what-history-teaches-us-about-todays-insider-threats/

PracticaleCommerce.com (2017). 105 Leading social networks worldwide. Retrieved from https://www.practicalecommerce.com/105-leading-social-networks-worldwide.

RapidMiner.com (2021). RapidMiner Studio. Comprehensive data science platform with visual workflow design and full automation. Retrieved from https://rapidminer.com/products/studio/

Roulin, N. (2016). Individual differences predicting impression management detection in job interviews. *Personnel Assessment and Decisions, 2*(1), 1.

Roulin, N., & Bourdage, J. S. (2017). Once an impression manager, always an impression manager? Antecedents of honest and deceptive impression management use and variability across multiple job interviews. *Frontiers in Psychology*, *8*, 29.

Safa, N. S., Von Solms, R., & Furnell, S. (2016). Information security policy compliance model in organizations. *Computers & Security*, *56*, 70-82.

Sahoo, G. (2013). Study of parametric performance evaluation of machine learning and statistical classifiers. *International Journal of Information Technology and Computer Science (IJITCS)*, *5*(6), 57.

Sari, D. F., Kurniawati, D., Prayitno, E., & Irfangi, I. (2019). Sentiment Analysis of Twitter Social Media to Online Transportation in Indonesia Using Naïve Bayes Classifier. In *Journal of International Conference Proceedings. 2*(1).

Schryen, G., Benlian, A., Rowe, F., Paré, G., Larsen, K. R., Gregor, S., & Petter, S. (2016). Standalone literature reviews in IS research: What can be learnt from the past and other fields? In *Proceedings of the International Conference on Information Systems (ICIS) in Dublin, Ireland*, 1-8.

Shapiro, A. H., Sudhof, M., & Wilson, D. (2018). Measuring news sentiment. Federal Reserve Bank of San Francisco.

Shaw, E. D. (2006). The role of behavioral research and profiling in malicious cyber insider investigations. *Digital Investigation*, *3*(1), 20-31.

Shaw, E. D., Ruby, K. G., & Post, J. M. (1998). The insider threat to information systems. *Security Awareness Bulletin*, *2*(98), 1-10.

Shi, Y., Booth, R. E., & Simon, J. (2017). The iterative effect of IT identity on employee cybersecurity compliance behaviors. *Association for Information Systems,* 1-5.

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. " O'Reilly Media, Inc.".

Silge, J., & Robinson, D. (2020). Welcome to Text Mining with R. *A Tidy Approach*. Retrieved from https://www.tidytextmining.com/sentiment.html

Simpson, W. R., & Foltz, K. E. (2017). Enterprise level security: insider threat counter-claims. In *Proceedings of the World Congress on Engineering and Computer Science, 1*, 1-6.

Soh, C., Yu, S., Narayanan, A., Duraisamy, S., & Chen, L. (2019). Employee profiling via aspect-based sentiment and network for insider threats detection. *Expert Systems with Applications*, *135*, 351-361.

Splunk.com (2020). Analyst Report. *The 2020 Magic Quadrant for SIEM.* Retrieved from https://www.splunk.com/en_us/form/gartner-siem-magic-quadrant.html

Spooner, D., Silowash, G., Costa, D., & Albrethsen, M. (2018). Navigating the insider threat tool landscape: Low cost technical solutions to jump start an insider threat program. In *Proceedings of the 2018 IEEE Security and Privacy Workshops.* 247-257.

Staar, P. W., Dolfi, M., Auer, C., & Bekas, C. (2018). Corpus Conversion Service: A machine learning platform to ingest documents at scale. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 774-782.

StackOverflow.com (2021). WEKA Weighted Average Question Mark, Retrieved from https://stackoverflow.com/questions/61725706/weka-weighted-average-question-mark

Tableau.com (2021). Get actionable insights fast. Retrieved from https://www.tableau.com/products/desktop

Talkwalker.com (2021). AI powered sentiment analysis puts your social mentions into context. Retrieved from https://www.talkwalker.com/artificial-intelligence

Terrell, S. R. (2015). *Writing a proposal for your dissertation: Guidelines and examples*. Guilford Publications.

The Mercury News (2019). Former Santa Clara swim club CEO pleads no contest to embezzlement. John Bitter illegally took nearly $500,000, and took out a loan in club's name to buy a bar in Arizona. Retrieved from https://www.mercurynews.com/2019/10/04/former-santa-clara-swim-club-ceo-pleads-no-contest-to-embezzlement/

Thornton, C., Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2012). Auto-weka: Automated selection and hyper-parameter optimization of classification algorithms. *CoRR, abs/1208.3719.*

Trifacta.com (2021). From Messy Files To Automated Analytics. Retrieved from https://www.trifacta.com

USCourts.gov (2017). Court interpreters deliver justice in all languages. Retrieved from https://www.uscourts.gov/news/2017/08/10/court-interpreters-deliver-justice-all-languages

USCourts.gov (2019). Public Access to Court Electronic Records. Retrieved from https://www.uscourts.gov/court-records/find-case-pacer

USCourts.gov (2020). U.S. District Court, California Northern District. PACER,

Case Management/Electronic Case Files. Retrieved from https://ecf.cand.uscourts.gov

USCourts.gov (2020). The United States District Court Northern District of Illinois. PACER, Case Management/Electronic Case Files. Retrieved from https://ecf.ilnd.uscourts.gov/

USCourts.gov (2020). California Southern District Court. PACER, Case Management / Electronic Case Files. Retrieved from https://ecf.casd.uscourts.gov/

Verhoeven, B., Daelemans, W., & Plank, B. (2016). Twisty: a multilingual twitter stylometry corpus for gender and personality profiling. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation,* 1632-1637.

Vicinitas (2020). Vicinitas helps track and analyze real-time and historical tweets of your social media campaigns and brands on Twitter. Retrieved from https://www.vicinitas.io

Vilendečić, B., Dejanović, R., & Ćurić, P. (2017). The Impact of Human Factors in the Implementation of SIEM Systems. *Journal of Electrical Engineering, 5,* 196-203.

Visualr.io (2021). Data Visualization and Analytics Platform. Retrieved from https://visualr.io

Wallace, R. J., & Loffi, J. M. (2014). The unmitigated insider threat to aviation (Part 2): an analysis of countermeasures. *Journal of Transportation Security*, *7*(4), 307-331.

Walters, W. H. (2016). Beyond use statistics: Recall, precision, and relevance in the assessment and management of academic libraries. *Journal of Librarianship and Information Science*, *48*(4), 340-352.

Webb, G. I. (2010). Naïve Bayes. *Encyclopedia of Machine Learning*, 713-714.

Williams, M. L., Levi, M., Burnap, P., & Gundur, R. V. (2018). Under the corporate radar: Examining insider business cybercrime victimization through an application of routine activities theory. *Deviant Behavior,* 1-13.

Yerby, J. (2013). Legal and ethical issues of employee monitoring. *Online Journal of Applied Knowledge Management*, *1*(2), 44-55.

Zaib, S., Asif, M., & Arooj, M. (2019). Development of Aggression Detection Technique in Social Media. *International Journal of Information Technology and Computer Science, 5*(8), 40-46.